ON THE DYNAMIC RELATIONSHIP OF PERCEPTUAL AND LEARNING AND MEMORY SYSTEMS: MECHANISMS FOR ADAPTIVE PERCEPTUAL CHANGE

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Lauren L. Emberson

January 2012

ON THE DYNAMIC RELATIONSHIP OF PERCEPTUAL AND LEARNING AND MEMORY SYSTEMS: MECHANISMS FOR ADAPTIVE PERCEPTUAL CHANGE

Lauren L. Emberson, Ph.D.

Cornell University 2012

Human cognition is supported, in part, by adaptive changes in perception that occur as a result of experience. This dissertation proposes a mechanism by which perception can adaptively change to reflect the structure of the environment. Environmental structure produces statistical regularities in sensory input. Starting very early in infancy and continuing across the lifespan, humans have the ability to pick up on these statistical regularities through a process called statistical learning. This dissertation presents empirical findings that statistical learning occurs through a dynamic and mutually-influential relationship between perceptual and learning and memory systems, where perceptual processes support and constrain learning, and learning and memory systems, in turn, shape future perception.

To this end, infants and adults participated in incidental learning paradigms. In these paradigms, participants experience sensory input endowed with novel environmental structure that can be learned through statistical regularities. After exposure, memory for the structure is tested. A number of experiments examined learning where the statistical information is held constant but the perceptual character of the information varied (e.g., whether the stimuli are auditory or visual or whether statistical information is based on perceptually varying exemplars).

Using these methods, Chapters 2 and 3 present evidence that auditory and visual perceptual processing biases what is learned from novel statistical regularities. Thus, perceptual processing can affect statistical learning even though statistical information is held constant. Chapter 4 presents evidence for a dynamic influence of perception on statistical learning throughout development. Specifically, the results in this chapter demonstrate that the relationship of auditory and visual learning is different in infant and adult learners and that auditory statistical learning changes in infants aged eight to ten months. Finally, Chapter 5 presents evidence that changes in perception can occur through experience with environmental structure and points to the involvement of learning and memory systems in this process. Experiments in this chapter find that variable yet regular exposure with a novel object supports changes in object perception. A combined eye-tracking/functional magnetic resonance imaging (fMRI) methods approach highlights the involvement of the medial temporal lobe, an important learning and memory system, in integrating across successive experiences to support changes in object perception.

Taken together, this dissertation presents empirical evidence that during experience with novel statistical regularities, perceptual processing affects learning, and learning and memory systems can affect perception. The interrelationship of perceptual and learning and memory systems could act as a dynamic mechanism supporting adaptive changes in perception across the lifespan. Implications for the fields of Developmental Psychology, Cognitive Psychology and Cognitive Neuroscience are explored.

BIOGRAPHICAL SKETCH

Lauren was born in Sarnia, Ontario, Canada to parents Marta Dutrizac and Chris Emberson and grew up in Metropolitan Toronto. At Dunbarton High School, she concentrated on studies in two broad areas: philosophy and literary criticism and the maths and sciences with a focus on biology.

In September 2001, she went on to undergraduate studies in the Faculty of Science at the University of British Columbia (UBC) in Vancouver. After completing the interdisciplinary Co-ordinated Science Program in her first year, she joined the new Cognitive Systems Program and declared a specialization in Brain and Behaviour. Again Lauren's academic pursuits encompassed both analytical philosophy and biology, this time merged in the study of human behavior in relation to neural systems. Her fourth year project, supervised by Dr. Lawrence Ward, was entitled "Somewhat Noisy Brains See Better: Prestimulus EEG dynamics and perceptual performance."

Lauren came to Cornell University in 2006 to start graduate studies working with Drs. Michael Goldstein, Michael Spivey, and Morten Christiansen. In 2008, she joined the Ithaca-Manhattan Graduate Initiative in Neuroscience (IMAGINE) Program and began research at the Sackler Institute of Developmental Psychobiology under Drs. Dima Amso, Jason Zevin and Bruce McCandliss. In 2010, Lauren became a visiting research fellow in the Cognitive, Linguistic, and Psychological Sciences Department of Brown University and continued her research with Dr. Amso. After the completion of her PhD, Lauren will join the Brain and Cognitive Sciences Department at the University of Rochester as a postdoctoral researcher under the supervision of Drs. Richard Aslin and Daphne Bavelier.

This document is dedicated to my mother, Marta Dutrizac Hanlon, and my high school biology and maths teacher, Michael Kramer. You have both inspired my creativity and aspiration for scientific knowledge but also emphasized the need for hard work and persistence along the way. All of these were necessary in abundance. Thank you.

ACKNOWLEDGEMENTS

I'd like to thank my academic mentors. First and foremost, the professors comprising my committee: Drs. Michael Goldstein, Dima Amso, Michael Spivey, Barbara Finlay and David Smith. Thank you for being immensely supportive, especially considering my unconventional graduate school trajectory. I cannot thank you enough for your guidance and encouragement, administered both individually and collectively. I'd also like to thank mentors who are not on my committee but who were also instrumental to the creation of this document: Drs. Jason Zevin, BJ Casey, Bruce McCandliss, Morten Christiansen, and Lawrence Ward.

Second, I'd like to acknowledge the financial and administrative support that I've received. I have been supported in part by a fellowship from the National Science and Engineering Counsel (NSERC) from the Government of Canada and the National Institutes of Health (NIH, R25 MH060478). My research with Drs. Amso and Zevin has been supported by the NIH (K01 MH07793 to DA, R01 DC007694 to BM, and R21 DC008969 to JZ). I am also very grateful for the excellent financial and administrative support that I have received from Cornell University. In particular, I'd like to thank the Cornell University Linkage Program for help with the numerous administrative and logistical hurdles involved in starting my graduate research at the Weill-Cornell Medical College in New York City.

Thank you to the families who participated in the experiments presented in this dissertation. Without your generosity, we would know much less about the wonders of development. I'd like to thank my fellow graduate students and wonderful research assistants without whom doing research would certainly be less enjoyable, interesting, and fruitful, with particular thanks to Rachel Albert

for all of her generous help in the B.A.B.Y. Lab and Dani Rubinstein for being a superstar research assistant.

Finally, I'd like to express my deep appreciation to my personal support network, without whom this document would simply not exist. This includes, but is not exclusive to, my immediate family: Marjorie Dutrizac, Michelle and Liam Hanlon, Chris and Madelaine Emberson and honorary family member, Cristina Selva; some wonderful friends: Catalina, Natasha, Talia; the baristas at Gimme! for not throwing me out. I'd like to thank Barbara and Tom Hale for imparting your collective wisdom during the last three years and John, thank you, truly, for all that I've learned during my time with you. Above all, I'd like to thank my mom, Marta Dutrizac Hanlon, for encouraging me to aspire from Day zero starting with the choice of my name.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Perception is where cognition and reality meet.

Neisser, 1976

Perception is the process of uncovering meaningful information about the world that is present in sensory input (J. J. Gibson, 1966; E. J. Gibson, 1969). Our sensory organs (e.g., eyes, ears) detect energy that comes from the environment (e.g., light reflecting off of a face, oscillations of air molecules emanating from a plucked guitar string). However, we do not perceive what lands on our sensory organs. For example, if we perceived the patterns of light that enter our eyes, our perception would change with every eye movement. Instead, we perceive the meaningful entities in the world: the objects, people, spatial layouts. The process of sensory transduction provides sensory input to perceptual systems which in turn engage in a process of differentiation, recognition and/or interpretation of this sensory input to arrive at a meaningful view of the external world.

While it seems natural for the Cognitive Scientist to approach perception as a kind of recording device, this dissertation adopts a theoretical framework in which vision is not like a camera, audition is not like an audio recorder, etc. Specifically, the author adopts an Ecological approach to perception, pioneered by J. J. and E. J. Gibson, which posits that perception "is an ongoing activity, a search for information .. and it provides us with fundamental knowledge that we take for granted" (E. J. Gibson & Pick, 2000, p. 3). While a camera takes discrete, independent snapshots, perception is ongoing and continuous, which entails that different moments of perception are not independent of each other

and cannot be considered in isolation. While a camera passively accepts visual input, perception is an *activity* and part of this activity is related to actively seeking out information about the world. While a camera takes a picture the same way regardless of what is in the picture, perception is biased by the information, or meaning, contained in the sensory input. In sum, while a camera passively records the sensory energy present at a particular point in space and time, perception is a temporally unfolding process and one that is closer to a type of exploration.

Moreover, perception is not a static process but one that is adaptively changing as a result of experience with the environment. Knowledge of the structure of the environment can facilitate a greater ability to pick out meaningful information in sensory input. Correspondingly, perceptual change can occur through a progressive incorporation of knowledge about the external world. Indeed, such experientially-dependent changes to perception are believed to support key perceptual-cognitive tasks such as language comprehension and face perception. Thus, changes in perception can be produced through experience with structure in the environment.

Even though ongoing research has identified many aspects of perception that change with experience, little is known about the *cognitive mechanisms* or *neural systems* that operate during experience to shape future perception. This dissertation is focused on the elucidating the nature of the mechanism(s) that translate experience into perceptual change. Specifically, the current work suggests that perception can be adaptively and progressively shaped through an interrelationship of perceptual systems with learning and memory systems, where an interrelationship is a mutual influence of two entities. Specifically,

perceptual and learning and memory systems can influence each other during experience with novel patterns or statistical regularities in sensory input. In this way, perceptual processing can ground what is learned from novel statistical regularities and the resulting knowledge or memories can shape future perception.

Learning and memory are interrelated concepts in cognitive psychology and cognitive neuroscience: Learning is a "relatively long-lasting change in behavior that is contingent upon the occurrence of a specific event or events" (Milgram, MacLeod, & Petit, 1987, p. 3); memory can be considered the consequence of learning and, through memory recall, past experience affects future cognition. Learning and memory can be "conceived as a fundamental *property* of brain systems and a natural *outcome* of the brain's various processing activities, rather than an *entity* in the brain." (Eichenbaum & Cohen, 2001, p. 11). For the purposes of this dissertation, the author does not distinguish between the specific contributions of learning and memory. For example, current studies investigating the nature of learning depend upon quantifying the resulting memories. Thus, learning and memory are collectively considered the process by which past experience can affect future behavior.

The current studies employ on a specific learning and memory task, statistical learning, and examine the role of perception in learning as well as the effects of learning on perception. Broadly, *statistical learning* is the acquisition of knowledge from incidental exposure to patterns or statistical regularities in sensory input. This type of learning has been implicated in the development of a number of perceptual-cognitive skills; many of these skills have are also supported by changes in perception. However, the role of perception in statistical

learning not been examined in depth and neither has the relationship between statistical learning and subsequent perceptual changes.

This dissertation argues for a dynamic and mutually-influential relationship, or interrelationship, between perceptual and learning and memory systems in the presence of novel statistical regularities in sensory input: Chapters 2 and 3 find evidence that perceptual processing both supports and constrains what is learned from novel statistical regularities; Chapter 4 presents evidence that the influence of perception in statistical learning is dynamic across stages of development; finally, Chapter 5 argues that learning from statistical regularities can directly change future perception.

1.1 Adaptive Changes in Perception

Experience supports lasting changes in perception. Effects of experience have been documented in a wide variety of perceptual phenomena, and notably in many phenomena that are relevant for daily life (i.e., ecologically-valid). Perceptual changes have been documented in the domains of speech perception (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Werker & Tees, 1984), face perception (Kelly et al., 2007; Nelson, 2010), cross-modal processing of faces and voices (Pons, Lewkowicz, Soto-Faraco, & Sebastián-Gallés, 2009), and spatial localization of sensory objects (Pagel, Heed, & Röder, 2009). In addition, many learning paradigms in the contemporary field of perceptual learning report changes in perception that are not ecologically-valid (e.g., increased sensitivity to a specific subthreshold sine-wave grating in a specific retinotopic region as demonstrated in Nishina, Seitz, Kawato, & Watanabe, 2007). These wide-ranging findings suggest that perceptual processing is malleable and changes

with experience.

Perceptual processing has been shown to be affected by a number of different types of life experience, including under conditions of altered experience such as sensory deprivation (e.g. deafness: Bavelier, Dye, & Hauser, 2006), type of language experience (Neville & Bavelier, 2002; Winawer et al., 2007) and even playing a musical instrument (crossed hand effect: Kóbor, Furedi, Kovács, Spence, & Vidnyánszky, 2006). The development of perceptual expertise has been shown to dramatically affect perception: Gauthier, Tarr and colleagues trained participants to categorize a novel class of visual object called 'Greebles' and found changes in visual perception (Gauthier & Tarr, 1997) as well as changes in neural activity of regions of the neocortex associated with face perception (Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999). Further research has established that these changes to visual perception are not specific to Greebles or to expertise acquired in the laboratory (Tanaka & Curran, 2001; Harel, Gilaie-Dotan, Malach, & Bentin, 2010).

Overall, the research reviewed above reveals a rich link between past experience and perceptual processing. Broadly, the ability for experience to affect perception is available in the first postnatal year of life (e.g. Kuhl et al., 1992; Nelson, 2010) and continues across the lifespan (e.g. Gauthier & Tarr, 1997). While there is evidence that some perceptual systems that are less permeable to experience-based change (see Neville, 2006), there has been increasing emphasis on the ability of experience to change perception in the field of cognitive psychology and cognitive neuroscience (e.g. Green & Bavelier, 2008).

1.2 Ecological Approaches to Learning and Development

Changes in perception will be considered within the context of Ecological approaches to learning and development. This theoretical approach emphasizes that learning can change what is perceived from sensory input, not simply what is responded to (E. J. Gibson, 1969; J. J. Gibson & Gibson, 1955; E. J. Gibson & Pick, 2000), and perception changes as a result of experience with the structure of the environment in the service of achieving a better fit between the external environment and an observer's perception of it.

In brief, Ecological approaches conceive of the world, existing independently of the observer, as structured and meaningful: this world has objects, people, events, etc. Moreover, the sensory input that an observer receives is filled with information about this richly structured environment (see J. J. Gibson, 1966); however, this information may not immediately available to the observer. Instead, with experience, the observer will have an increased ability to pick out meaningful and relevant information from sensory input. Such a process is the essence of perceptual learning and development, as outlined by E. J. Gibson (1969).

An intuitive example of perceptual learning is wine tasting. Before acquiring any expertise, one has a rich sensory experience of wine. After learning about the red wines of Bordeaux and the Rhône, one still receives the same sensory input. However, this prior experience changes one's ability to pick out relevant information from it. For example, one is able to detect the presence of the Syrah grape when experiencing wine from the Rhône while the wine from Bordeaux includes information about the Cabernet Franc grape. In this way, the sensory

information remains the same but experience changes what information the observer is able to pull out of sensory input.

Similarly, ecologically-valid changes of perception, as reviewed above, can result from changes in the ability to pick out information from sensory input. For example, infants' ability to pick out speaker identity is influenced by language experience. Specifically, 7-month-old infants are unable to detect changes of speaker in an unfamiliar language but are able to do so for familiar languages (E. Johnson, Westrek, Nazzi, & Cutler, 2011). [1] Thus, in the second half of the first post-natal year, infants' ability to pick out identity in sensory information about voices is influenced by experience with the sensory input. Taking evidence from a very different domain, the crossed hand effect is a phenomenon where perceptual processing, even of non-tactile stimuli, is biased when participants' hands are crossed in space. Recent work on the development of this effect has found that early in childhood, there is no difference in processing of tactile input between crossed and uncrossed positions. However, between the ages of three and five, children become increasingly better at processing tactile stimuli in the uncrossed position (the position that they use most often when interacting with their world) with no changes in processing in the crossed po-

[1] The visual analogue to this effect could be considered to be the other-race effect. The other-race effect, reviewed above, is an example of perceptual narrowing: the other-race effect is the decreased sensitivity to the individual identity of faces from non-familiar racial groups. This effect begins to develop by six months of age (Kelly et al., 2007) and is affected by experience (Pascalis et al., 2005). Pioneering work on perceptually narrowing was done by Kuhl, Lewkowicz, and their colleagues (e.g., Kuhl et al., 1992; Kuhl & Rivera-Gaxiola, 2008; Lewkowicz & Ghazanfar, 2009; Pons et al., 2009). Given that E. Johnson et al. (2011) are the first to document a difference in talker identification as a result of language experience, and only in a single age group, it is not clear whether this effect is also the result of perceptual narrowing. Relatedly, it is interesting to consider whether perceptual narrowing could also be an example of increases in the ability to pick out information in sensory input. Perceptual narrowing is often considered to be a by-product of the process of neural specialization that arise through experience and has been proposed to support increased perceptual abilities in the areas of specialization to the detriment of processing for stimuli that are not experienced (see Kuhl & Rivera-Gaxiola, 2008). In accord with this view, perceptual narrowing could be considered the by-product of an increased ability to pick out information in sensory input.

sition (Pagel et al., 2009). Thus, the crossed hand effect can be considered to be a result of an increased ability to pick out sensory information when hands are in an uncrossed position. In sum, taking examples from two very different domains, the author argues that changes in perceptual abilities during development could be considered as differences in the ability to pick out information in sensory input.

1.2.1 Adaptive Fit or Learning through a Reduction of Uncertainty

Why do changes in perception occur as a result of experience? According to the Ecological approach, perceptual changes occur to support a more harmonious fit between the observer, specifically their perception, and their environment or niche (J. J. Gibson, 1966; E. J. Gibson, 1969). Similarly, many contemporary views characterize learning as a progressive reduction of uncertainty between internal predictions and the outcomes of events (e.g. McClelland, 2002; Schultz, 2006). These two views are essentially complementary to each other as the best fit between the organism and its niche would result in the least possible prediction error.

Effective perception is a key aspect of this process. According to E.J. Gibson, "[p]erceptual activity searches the environment in the service of adaptive, economical action. Locating the path perceptually saves one from stumbling. Perceiving the cliff edge can save a climb up again or even save life" (E. J. Gibson, 1969, p. 121). Thus, changes in perception serve to support a better organism-niche fit and can be conceptualized as occurring through a reduction of uncer-

tainty.

1.2.2 Experience in a Structured Environment

As summarized above, sensory input contains information about the organization or structure environment. In this way, sensory input could be said to be *structured*. An alternative theoretical view would hold that sensory input is not rich with information about the structure of the world either through a lack of coherent structure in the environment or the inability of sensory input to convey this information. If sensory information was not structured (e.g., a learner was presented with random or unstructured sensory input), there would likely be no changes to perception as a result of experience because there would be no method for increasing the fit of perception or decreasing uncertainty. Thus, experience with structure present in sensory input is necessary to support changes in perception.

In sum, Ecological approaches postulate that experience with structure or meaningful information in sensory input can influence perception. While the goal of development may be "to learn structure through space and time" (Goldstein et al., 2010, p. 249), this section suggests that learning of environmental structure can occur incidentally through an organisms pursuit of a better fit with its ever changing niche through reducing uncertainty about sensory input. These changes occur with an observer's increased ability to pick out relevant information in sensory input.

1.3 What is Statistical Learning?

Through experience, the structure of the environment can increasingly influence perception. This dissertation aims to investigate the nature of the mechanisms that support experience-based perceptual change. To this end, the studies presented focus on the relationship of a specific learning paradigm–statistical learning–and perception. This section presents a brief survey of the field of statistical learning, the phenomena that it strives to explain and the paradigms it employs. This section begins with a brief examination of statistical learning in relation to the theoretical framework set out in the previous section.

1.3.1 Ecological Statistical Learning

Broadly, statistical learning is the process of gaining knowledge of the structure of the world through incidental experience with statistical regularities in sensory input. Building from the Ecological approaches presented above, one way that the structure in the environment can be instantiated in sensory input is through the presence of statistical patterns. For example, take the phrase "pretty baby" which would typically be produced as a continuous utterance /prI'tibeI'bi/. For an infant exposed to English, this phrase is a part of the structure of the ambient language. The transitional probabilities, the probability that one will perceive Y given the perception of X, as well as other statistical regularities (e.g. co-occurrence frequency) are higher between syllables that cohere to form a word in this phrase (e.g. "pre" and "ty") than syllables that cross word boundaries (e.g. "ty" and "ba"). Numerous studies have demonstrated that infants as young as 8-months-old and adults can use statistical information

defined across syllables to segment a continuous speech stream (Saffran, Aslin, & Newport, 1996). In other words, an observer can learn to pick out meaningful information through experience with statistical patterns in sensory input.

The detection of statistical regularities in their sensory input is a fruitful avenue to reducing uncertainty or prediction error and to learn about the structure of the environment. Returning to the example of the phrase ”pretty baby,” identification of ”pret” as the first syllable in the word ”pretty” will reduce uncertainty of what will be heard next by predicting that one will hear ”ty.” Thus, one of the methods for effective reduction of prediction error is to identify patterns in input that can be used to predict future events. Indeed, infants and children have been found to exploit the statistical patterns present in their environment for more efficient or rapid processing of sensory input. Relevant to the previous example, infants as young as 18 months can use the first half of a spoken word to pick out its visual referent as quickly as infants who hear the full word (Fernald, Swingley, & Pinto, 2001). Similarly, children have also been found to use markers of grammatical gender, spoken preceding words, (e.g., la vs. el in Spanish) to identify visual referents (Lew-Williams & Fernald, 2007). Thus, the identification of patterns found in sensory input is one important way that organisms can learn about the structure of the world and thus reduce uncertainty.

It has also been argued that observers engage in predictions based on statistical regularities even when they do not have control over their sensory input. While many studies have found that observers across the lifespan demonstrate changes in behavior as a result of experience with statistical regularities in sensory input. Few studies have examined the learning mechanisms by which this occurs and specifically whether learning in these behaviorally passive tasks oc-

curs through a reduction of prediction error and/or a generation of predictions. However, Haith, Hazan, and Goodman (1988) found that 3.5-month-old infants do develop predictions simply as a result of passive experience with a sequence of visual input. Infants were presented with alternating visual scenes. During exposure, infants exhibited more anticipatory eye movements (produced before visual presentation) but also faster, non-anticipatory, eye movements (produced in response to visual presentation) as compared to infants in a control group that viewed similar scenes in an unpredictable order. The authors ask "[w]hy would a baby gratuitously develop expectations for series of events over which it has no control?" The infants could simply respond *ad hoc* or reflexively to the stimuli. While this is still an open question, Haith et al. (1988) argue that the development of expectations allows the infant to bring some aspect of an uncontrollable sequence under their control. The author argues that the development of these expectations may be akin to building an internal model of the external environment that can in turn shape cognitive processes such as eye movements. Turk-Browne, Scholl, Johnson, and Chun (2010) also argues that passive experiences with statistical regularities results in neural expectations or predictions for successive visual stimuli (see also E. J. Gibson, 1969, for an argument for an endogeneous motivation to reduce uncertainty even as a result of passive experience). Thus, while there are few studies that have explicitly examined the mechanisms of incidental learning through passive experience with statistical regularities, there is some evidence that statistical learning also occurs generation of predictions and a reduction of prediction error (see Section 6.2.1 for a more in depth discussion of this topic).

In sum, statistical learning is the process of gaining knowledge of the structure of the world through incidental experience with statistical regularities in

sensory input. Statistical learning is one method for an observer to learn to pick out meaningful information in sensory input and reduce prediction error.

1.3.2 Language-Inspired Statistics

While there is a potentially vast range of environmental statistics that an organism could encounter and learn from, the literature which identifies itself with the term "statistical learning" focuses on a small subset of these statistics. This field, and correspondingly the statistics they are concerned with, are inspired by the problem of language learning.

Part of what makes the acquisition of language a compelling developmental task from the perspective of learning from statistics, is that there are many different "levels" of statistical information in a single speech stream; there are statistical regularities corresponding to phonological, lexical and grammatical levels of language processing. In fact, the relevance of statistical learning has been demonstrated for many of the important developmental aspects in language acquisition across levels of processing. Table 1.1 presents a list of different language learning tasks and the "level" of statistical information believed to underlie that part of language learning. Below, some of these language learning tasks are explained in greater detail:

- *Acquisition of phonological categories*: This language learning task is believed to, in part, be facilitated through exposure to exemplars in different distributions along acoustic dimensions (e.g. Voice Onset Time). Exposure to a bimodal distribution can reflect a language which makes a phonological distinction within this dimension whereas a unimodal distribution likely

Table 1.1: Examples of Statistical, Language Learning Tasks and Their Corresponding "Level" of Learning in the Speech Stream

Learning Task	Level	Sample Study
Phonological categories	Phonological	Maye, Werker, and Gerken (2002)
Phonological pattern (e.g. VCV)	Phonological	Goldstein and Schwade (2008)
Word segmentation	Lexical	Saffran et al. (1996)
Word–visual referent association	Lexical	Yu and Smith (2007)
Artificial grammar learning	Syntactic	Gomez and Gerken (1999)
Sentence frames	Syntactic	Fernald and Hurtado (2006)

reflects that the language does not make such a distinction (Clayards, Tanenhaus, Aslin, & Jacobs, 2008; Maye et al., 2002; Yoshida, Pons, Maye, & Werker, 2010).

- *Learning the characteristic phonology of the ambient language through social interaction*: A statistical learning task pioneered by Goldstein and Schwade (2008) combines exposure to a variable phonological pattern with social feedback. When varying exemplars of this novel phonological pattern are presented contingently with infant babbling, infants will exhibit novel production of this pattern.

- *Segmentation of words from a continuous speech stream*: Co-occurence frequencies and transitional probabilities of successive sounds in a continuous speech stream facilitate the segmentation of words. For example, if a speech stream includes novel words such as "dakpel" and "rusjux," the learner can use the higher-order statistical information to differentiate within vs. between word sequences of syllables (Saffran et al., 1996). This is the most common statistical learning task and has been demonstrated with a wide range of stimuli. Novel variants on this task will be employed in Chapters 2, 3, and 4.

- *Artificial grammar learning (AGL)*: A finite-state grammar is designed to produce a large number of possible sequences of novel words. For example, a sequence that is produceable from the finite-state grammar illustrated in Figure 1.1 would be "PEL-TAM-RUD-RUD" whereas a sequence like "RUD-PEL-VOT-TAM" would not be permitted by the grammar. The learner is asked to discriminate between permissible and non-permissible sequences after exposure to a number of different sequences from this grammar. This task is aimed at examining a higher-level abstraction for sequence learning reflecting syntactic processing (Gomez & Gerken, 1999).

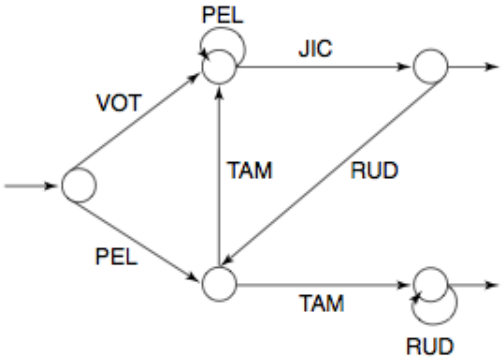


Figure 1.1: Sample artificial grammar for a statistical learning task by Gomez and Gerken (1999)

Likely because the field of statistical learning is focused on the developmental task of language acquisition, even statistical learning experiments that do not directly relate to language learning employ similar language-inspired statistical patterns. Most saliently, the statistical information is presented sequentially in a stream and employs similar types of statistical regularities (e.g., transitional probabilities between sensory objects).

Even though the sequential statistics explored by the field are inspired by language, statistical learning is not a skill that is specific to the domain of lan-

guage and thus is domain-general. Many studies have found learning using non-linguistic stimuli. For example, the statistical learning task of (Saffran et al., 1996) has been replicated using non-linguistic auditory stimuli (e.g. Saffran, Johnson, Aslin, & Newport, 1999) and a number of types of visual stimuli (Brady & Oliva, 2008; Kirkham, Slemmer, & Johnson, 2002). Statistical learning has also been established using spatial-arrayed stimuli presented simultaneously (Fiser & Aslin, 2001, 2002; Saffran, 2002), extending evidence of statistical learning from sequential presentation of visual and auditory stimuli. This work is theoretically vital for establishing statistical learning as a domain-general contribution to language learning and opening up the possibility for contributions of SL to development in other domains. This will be expanded upon in Chapters 3, and 4.

1.3.3 What Is Statistical Learning: Conclusions

Essentially, statistical learning is the process of gaining knowledge of the structure of the world through experience with statistical regularities in sensory input. Consistent with Ecological theories of learning and development, the use of these regularities can increase the accord between an observer and their environmental or developmental niche.

The field of statistical learning is largely coalesced around the important developmental task of learning an ambient language. As a result, the paradigms of the statistical learning literature are language-inspired. This is most evident in the use of sequential statistics in statistical learning tasks. Despite a grounding in language, the field of statistical learning is domain-general in focus, as sta-

tistical learning has been demonstrated with many different sources of sensory input, suggesting a role for statistical learning in many different developmental tasks.

1.4 What Isn't Statistical Learning?

As outlined above, the widely-accepted definition of statistical learning is a sensitivity to statistical information in sensory input (Conway & Christiansen, 2006; Gómez, 2006; Perruchet & Pacton, 2006; Saffran & Sahni, in press). This definition encompasses the wide range of phenomena that have been studied under the term "statistical learning," such as learning from the distribution of exemplars in a phonological category and using transitional probabilities defined across syllables to segment words in a speech stream. However, this definition also easily engulfs many other types of learning and memory. In this section, we consider what *isn't* statistical learning and how statistical learning relates to other forms of learning. We begin with a brief discussion of the view of learning and memory taken in this dissertation.

1.4.1 Multiple Learning and Memory Systems as Computational Niches

Learning and memory is not a singular behavioral phenomenon, nor is it mediated by a single brain region, neural system or cognitive mechanism. It is commonly accepted that learning and memory comprise multiple systems and constructs. However, there is no consensus on how to characterize the multiple

types of learning and memory.

Arising from memory research in the second half of the 20th century, the most well-known and influential characterizations of learning and memory are based on crisp psychological and corresponding neural dissociations. The *Declarative Memory* hypothesis makes a distinction between declarative and procedural (non-declarative) memory (see Squire, Stark, & Clark, 2004, for a recent review). Figure 1.2 presents a visual depiction of the major divisions of learning and memory systems according to this hypothesis. Declarative memory is conscious and detailed; semantic knowledge, such as consciously remembering a dissertation defense date, is the classic example of declarative memory. This type of memory is contrasted with procedural skills, such as visuomotor tasks, that cannot be verbally articulated and operate outside conscious awareness. According to the Declarative Memory hypothesis, these two forms of memory are supported by distinct neural systems: Declarative memory is supported by the medial temporal lobe (MTL) and non-declarative memory is supported by the basal ganglia (BG), possibly in combination with other neural structures. The dichotomous characterization between declarative and non-declarative memory is often employed analogously to the explicit–implicit learning distinction (with and without conscious awareness, respectively; e.g. Keele, Ivry, Mayr, Hazeltine, & Heuer, 2003; Nomura & Reber, 2008). A similar characterization based in part upon a combination of these two psychological characteristics comes from Schacter and Tulving (1994) and characterizes five major types of memory based on psychological characteristics.

However, the theory of distinct declarative and procedural systems of learning and memory has increasingly come under fire. Notably, a number of recent

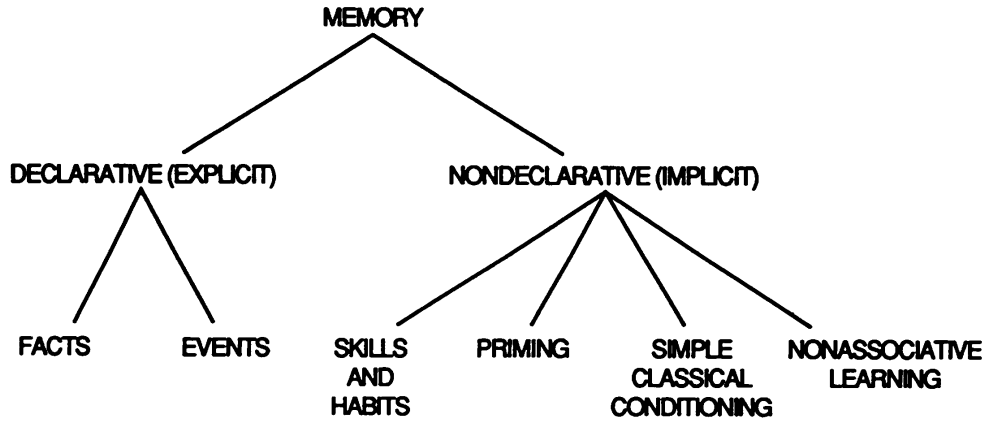
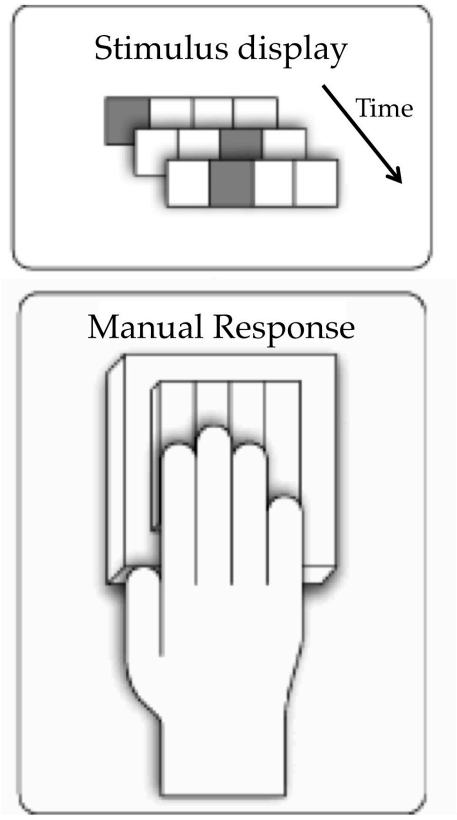


Figure 1.2: Division of multiple learning and memory systems as proposed by Squire and Zola-Morgan (1991)

studies have found the involvement of the MTL in non-declarative learning and memory tasks (e.g. Chun & Phelps, 1999; Turk-Browne, Yi, & Chun, 2006). Relevant to the current series of studies, the MTL has been implicated in implicit learning of sequences (e.g. Schendan, Searl, Melrose, & Stern, 2003). Curran (1997) examined performance of MTL amnesiac patients on the serial reaction time task, SRTT, a well-known example of procedural memory believed to be mediated by the BG. In this task, visual stimuli with distinct locations are presented, and participants are simply asked to press a corresponding response key as quickly as possible. Unbeknownst to them, the stimuli are presented in a predictable sequence. Previous studies had found spared performance in the SRTT by MTL amnesiacs supporting the canonical divisions of learning and memory. However, Curran (1997) found that when sequences' statistical properties were controlled such that they could only be learned by second-order associations, MTL amnesiacs show impaired performance. The first- and second-order sequences employed by Curran (1997) are presented in Figure 1.3. This finding demonstrated that modifications of the computational demands of the learning task, not the psychological character of the task, changed the mediating neural

systems.



First-order Sequence

A-B-A-D-B-C-D-C-A-D-B-C

Each element, e.g. A, is followed by another element (e.g. D) 67% of the time and by a second element 33% of the time (e.g. B)

Second-order Sequence

A-B-A-D-B-C-D-A-C-B-D-C

Each element is equally likely to be followed by any other element and can only be predicted based on a combination of previous stimuli (e.g. A-B predicts A while D-B predicts C)

Figure 1.3: Illustrations of the Serial Reaction Time Task (SRTT, adapted from Clark & Ivry, 2010) with sequences employed by Curran (1997)

An alternative characterization of multiple learning and memory systems is proposed by O'Reilly and colleagues (see Atallah, Frank, & O'Reilly, 2004; McClelland, McNaughton, & O'Reilly, 1995; O'Reilly, 2006). The core idea of this approach is that "different brain areas are specialized to satisfy fundamental tradeoffs in the way that neural systems perform different kinds of learning and memory tasks" (Atallah et al., 2004, p. 253). As such, this view does not propose psychological or neural dichotomies between types learning and memory. Instead, the view presented by O'Reilly and colleagues characterizes different learning and memory systems according to their neurocomputational proper-

ties. The diversity of computational properties of the MTL, the basal ganglia (BG) and other neural regions involved in learning and memory tasks (e.g. the prefrontal cortex, PFC), can be described as complementary in that different learning and memory tasks require different types of computation.

O'Reilly and colleagues propose viewing the relationship between types of learning and memory tasks and their mediating neural systems in relation to their neurocomputational properties. This entails that learning and memory tasks can differentially engage neural systems based on their computational demands as exemplified by Curran (1997). Second, this characterization highlights the possibility that multiple systems, previously viewed to be distinct, dissociable or competing (Poldrack & Packard, 2003), could all contribute to a given learning and memory task, albeit to different aspects of the task. Thus, multiple learning and memory systems can be viewed as computationally complementary.

1.4.2 Computational Demands of Statistical Learning Tasks

Holding the view that the engagement of neural systems can be characterized based on computational characteristics of a task places focus on the *types* of computational demands of a given task rather than the psychological character of the task (e.g. procedural vs. declarative). Correspondingly, in order to answer the question, "What isn't statistical learning?" we must consider the computational demands of statistical learning as they relate to other learning tasks.

To this end, canonical learning and memory tasks will be discussed in relation to their intended learning outcomes. For example, episodic memory tasks

designed for human participants, typically present multiple sources of information to remember such as a picture presented in a specific color (Ghetti & Angelini, 2008). The intended information to be learned and remembered is the conjunction of these sources of information and/or one part of the stimulus (e.g. the identity of the picture). However, participants could encode any number of memories based on this task (e.g. the order of trials). The current section will only discuss the computational aspects of the intended task outcomes.

An important characteristic of statistical learning studies is that the information to be learned is instantiated over many individual experiences. A statistical pattern is not evident in any single sensory object or event but is characterized through successive events. This is true both for sequential statistical learning tasks focused on segmentation (e.g. Saffran et al., 1996) and for distributional statistical learning tasks (Maye et al., 2002; Yoshida et al., 2010). Thus, in statistical learning tasks, the information intended to be learned is highly distributed over many successive stimuli or experiences.

There are a number of learning and memory tasks where the information intended to be learned is presented in a very few number of trials and sometimes a single trial: This is the case for perceptual priming tasks where recently encountered words or pictures influence subsequent performance on recognition or completion tasks (e.g. Graf & Schacter, 1985) and for episodic memory tasks as described above. Considering the distribution of the to-be-learned information provides the first answer to what isn't statistical learning: If learning task could be completed in a single or very few number of trials, it *isn't* a statistical learning task.

Figure 1.4 presents a graph illustrating the position of a number of learning

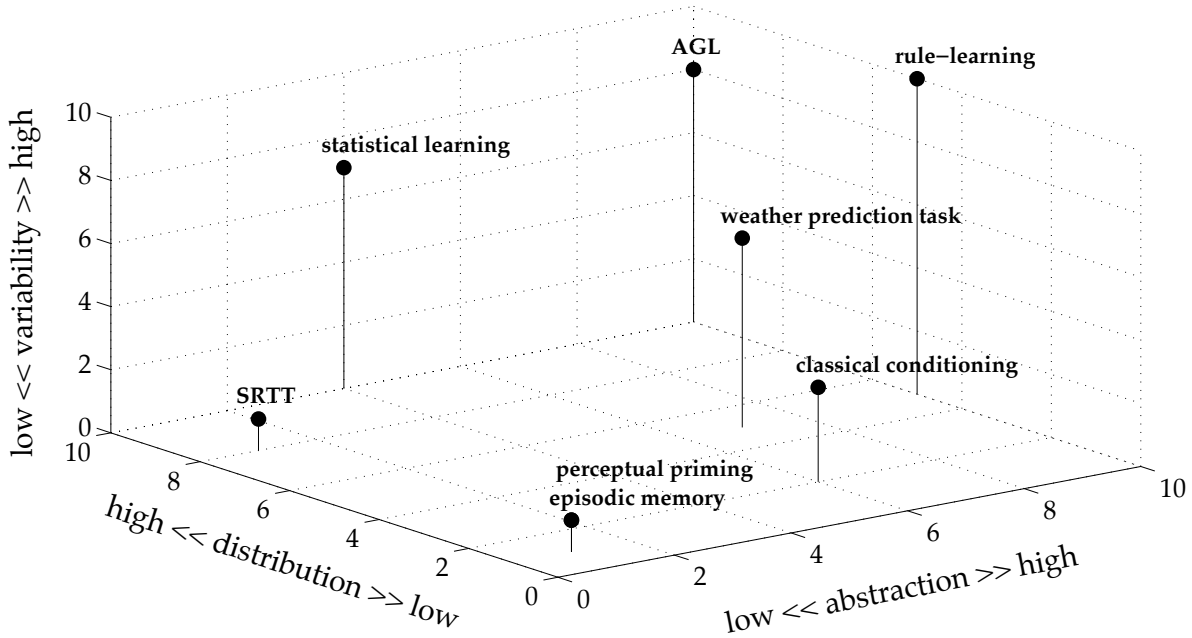


Figure 1.4: Learning and memory tasks presented in three dimensions of computational demands: distribution (x axis), variability (y axis) and abstraction (z axis). Inspired by: Spivey and Knowlton (2008)

and memory tasks along the computational dimensions to be discussed in this section. The x-axis (closest and to the left) represents the *distibution* of to-be-learned information in a given task. As discussed above, perceptual priming and episodic memory are very low in the dimension of distribution (forward and right) while tasks like the SRTT and statistical learning are very high on the dimension of distribution (back and left).

A second defining characteristic of statistical learning tasks is the *variability* of exposure. Statistical learning tasks have a relatively high degree of variability: In a segmentation task, the different "words" to be learned are presented in random order. This effectively reduces the higher-order statistics of between word syllables while reinforcing within-word statistics. Similarity, in a distri-

butional statistical learning task, there is a high degree of perceptual variability (e.g. eight different exemplars for a given category used in Maye et al., 2002; Yoshida et al., 2010). Moreover, a specific form of statistical learning task (non-adjacent statistical learning) has been found to rely on high variability (Onnis, Monaghan, Christiansen, & Chater, 2004).

There are other learning and memory tasks that present their stimuli with a high degree of variability. These are tasks where participants are expected to learn a set of "rules" or categories such as the aptly named rule-learning task (Marcus, Vijayan, Rao, & Vishton, 1999; Saffran, Pollak, Seibel, & Shkolnik, 2007), artificial grammar learning (AGL), and weather prediction task (Shohamy, Myers, Onlaor, & Gluck, 2004).

The amount of variability in a task is related to the distribution of information in so far as tasks with very low distribution simply cannot allow for variability (e.g. perceptual priming). However, consider the SRTT where distributed information (e.g. a sequence of ten trials) is simply repeated without variation (e.g. Nissen & Bullemer, 1987). Thus, a second dimension that distinguishes statistical learning from some other learning tasks is the degree of variability during exposure (see Green & Bavelier, 2008, for a discussion of the role of variability in procedural learning tasks and behavioral plasticity). The dimension of variability is presented along the y-axis in Figure 1.4.

A third dimension which distinguishes SL tasks is *abstraction*, which refers to the level of fidelity of the to-be-learned information. At a level of high abstraction, there is low fidelity between what a participant is exposed to and the to-be-learned information and vice versa. Recall the SRTT employed by Curran (1997), where participants could only learn the sequence based on second-order

statistics. This task has a higher degree of abstraction than an SRTT such as the version employed by Nissen and Bullemer (1987) where participants were intended to simply learn the succession of the entire sequence.[2] A task with a very low degree of abstraction is an episodic memory task which requires a very precise recall between exposure and test. Tasks with a very high degree of abstraction are rule-learning tasks (AGL, rule-learning).

Tasks that can be described as statistical learning have an intermediate degree of fidelity. Statistical learning does not require participants to learn the actual order of the sequence as in SRTT tasks. Instead they are asked to find the probabilistic patterns in the input which requires a greater degree of abstraction. However, statistical learning tasks are distinct from tasks where the participant is expected to generalize as in a rule-learning task (However, see Brady & Oliva, 2008; Lany & Saffran, 2010, which present a higher degree of both variability and abstraction than the typical statistical learning task). A distinction between the mechanisms of rule-learning and statistical learning has been suggested, but their intersection, if any, is still an open question (See Marcus et al., 1999; Seidenberg, MacDonald, & Saffran, 2002). Thus, examining the dimension of abstraction presents three groups of learning tasks: those with high abstraction that exemplify rule-learning tasks, those that require a low degree of abstraction like the canonical SRTT, episodic memory and perceptual priming and tasks that require an intermediate degree of abstraction that best fit the definition of statistical learning. Abstraction is presented along the z axis in Figure 1.4. [3]

[2]As pointed out by Curran (1997), participants likely learned the sequence in Nissen and Bullemer (1987) using simple two-response associations which would be a lower level of abstraction than the second order sequence employed by Curran (1997) and a higher level of abstraction than intended by the authors.

[3]There is a positive relationship between the dimensions of abstraction and variability: Tasks that tend to be high in variability also tend to be high in abstraction. This relationship is dis-

This section considered statistical learning–learning from statistical regularities presented in sensory input–and other learning tasks along three dimensions. Statistical learning cannot occur in a single trial like perceptual priming or episodic memory. Instead, a statistical pattern must be presented over many trials. Thus, statistical learning has a greater distribution than other learning tasks. However, unlike the SRTT, statistical learning is not simply a memorization of a sequence of stimuli. Thus statistical learning tasks have a higher degree of variability and a higher degree of abstraction. Statistical learning can also be dissociated from tasks more focused on learning an abstract rule based on sensory input. These tasks require a higher degree of abstraction and often also have a higher degree of variability. However, tasks such as classical conditioning occupy a similar space in this multidimensional space as statistical learning tasks do.

1.4.3 Statistical Learning As Theoretical Point of View

This current section has been considering what *isn't* statistical learning. This section is necessary because the definition of statistical learning can be seen to encompass other types of learning. The previous section argued that tasks that are considered statistical learning can be dissociated from many other learning tasks along a number of computational dimensions, such as variability and abstraction. This section explores an alternate, if complementary, way to distinguish statistical learning from other forms of learning: The field of statistical learning embodies a theoretical point of view. Specifically, this section examines the differences in theoretical motivations for classical conditioning and statisti-

cussed in more detail in Chapter 2.

cal learning.

Consider these two definitions of classical conditioning:

> Conditioning refers to the acquisition of a reflexive response to stimuli that are repeatedly paired with an unconditioned stimulus that elicits the reflex, following the Pavolvian protocol. Eichenbaum & Cohen, 2001, p. 155

The second definition is from Rovee-Collier (1986): "[Classical conditioning is] a process by which organisms acquire predictive information about the structure of their environments" (p. 140). These definitions notably diverge on the theoretical focus the authors are bringing to the definition. Eichenbaum and Cohen focus on the psychological dissociations between types of learning and memory, while Rovee-Collier approaches questions of learning and memory from an Ecological perspective emphasizing the interrelationship of an organism and its environment and the role of learning in development.

Both of these definitions can be considered for how they relate to the definition of statistical learning: The former definition of classical conditioning does not overlap with the definition of statistical learning. It emphasizes the role of the changing reflexive response in classical conditioning and the importance of both conditioned and unconditioned stimuli. None of these aspects of classical conditioning are included in the paradigms comprising the field of statistical learning. However, the latter definition emphasizes the importance of predictability in learning and the function of learning for an organism and in this way, it overlaps with the definition of statistical learning. In fact, the definition of classical conditioning presented by Rovee-Collier (1986) could be a substitute

definition for statistical learning.

The field of statistical learning is borne from the assertion that there is enough structure present in sensory input to support development, specifically, the development of language abilities. In this way, the field of statistical learning is in accord with an Ecological perspective. This is discussed in greater detail in Section 1.3.1.

Essentially, there is a convergence between the definition of classical conditioning presented by Rovee-Collier and statistical learning within an Ecological perspective. Such a convergence could be also envisioned for learning tasks such as the SRTT and perceptual priming. Uniting different learning and memory tasks through their theoretical focus provides an important synergy between interrelated related learning and memory processes even though they may be dissociable on the dimensions of *distribution, variability,* and *abstraction*.

1.5 The Interrelationship of Perception and Learning and Memory during Statistical Learning

This dissertation argues that perceptual and learning and memory systems influence each other during experience with novel statistical regularities. An influence of learning and memory on perception would allow knowledge gained through statistical learning to shape future perception. An influence from perceptual processing on learning and memory would ground knowledge, gained through exposure to statistical regularities, in current perceptual processes.

This section asserts that a grounding of statistical learning follows from the

theoretical framework for this dissertation, provided by Ecological approaches to perception and development. The implications are explored, starting with considering statistical learning relative to an observer's niche. Conceptualizing learning as being grounded provides a novel proposal for the nature of the mechanisms supporting statistical learning, and this section begins with a review of the current understanding of the mechanisms supporting statistical learning.

1.5.1 Current Views on the Mechanisms of Statistical Learning

Statistical learning is commonly believed to be supported by a largely abstract learning mechanism focused on the statistical information in the sensory input. There have also been suggestions that statistical learning is developmentally invariant, and in this way cognitively separated from developmental changes in perception, for example.

Early statistical learning studies focused on establishing the domain-generality of this learning ability. Seminal papers established that SL is not limited to language input (e.g. Saffran et al., 1999) or sequential statistics (e.g. Fiser & Aslin, 2001, 2002). Some studies argued that not only can learning proceed with non-linguistic input but that learning can be equivalent across input types (Saffran et al., 1999). As discussed in Section 1.3.2, this work was theoretically vital for establishing statistical learning as a domain-general mechanism that supports language learning. The establishment of domain-generality also introduced the possibility that statistical learning could contribute to the development in other perceptual-cognitive domains.

Domain-generality is has been a prominent influence on the investigation of the cognitive mechanisms underlying statistical learning. Many influential models and theories have presupposed a mechanism that treats all types of input stimuli (e.g., tones, shapes, syllables) as equivalent beyond the statistical structure of the input itself (e.g., Altmann, Dienes, & Goode, 1995; Perruchet & Pacton, 2006; Reber, 1989; Shanks, Johnstone, & Staggs, 1997). Thus, statistical learning is commonly seen to be supported by an abstract learning system focused on statistics exclusively.

Some studies have also suggested that the mechanisms supporting statistical learning are invariant across developmental time: Saffran, Newport, Aslin, Tunick, and Barrueco (1997) compared auditory SL between adults and children and Kirkham et al. (2002) examined visual SL in infants from two to eight-months-old; both studies found no difference in learning outcomes across age groups. If statistical learning is developmentally invariant, then it is separate from changes of perception across developmental time. However, statistical learning studies typically employ stimuli that are novel to the observer (e.g., novel words consistent with the phonology of the ambient language) that may not be perceived differently as no observer (an infant or an adult) has had prior experience with them.

Building from the importance of domain-generality, there has been recent interest in the constraints on statistical learning by other cognitive factors such as attention. Early studies argued that statistical learning could occur outside of selective attention, the process that biases cognitive processing to specific aspects of sensory information (Saffran et al., 1999). However, recent work has suggested that selective attention does modify statistical learning and in

fact, may be essential for learning to occur (Toro, Sinnett, & Soto-Faraco, 2005; Turk-Browne, Jungé, & Scholl, 2005). It has also been suggested that perceptual grouping and attention may interact when learning from statistical regularities (Baker, Olson, & Behrmann, 2004; Pacton & Perruchet, 2008). Beyond establishing attention as an important factor in learning, this work provides some initial evidence that statistical learning is not unaffected by other cognitive processes.

Thus, the understanding of statistical learning has progressed from being characterized as language-specific, to domain-general and abstract. Some current thinking has emphasized the effects of cognitive constraints such as attention.

1.5.2 Statistical Learning As Grounded in the Observer's Niche

This dissertation diverges from the domain-general view of statistical learning reviewed above. Domain-general views imply that statistical learning is affected mainly, or exclusively, by the presence of statistical information in sensory input. The Ecological framework employed for this dissertation emphasizes that what is learned must be grounded in the niche in which the observer exists (E. J. Gibson, 1969; E. J. Gibson & Pick, 2000; Rovee-Collier & Cuevas, 2009). A *niche* is the relationship between an observer's cognitive system and their environment. An observer's current niche is affected by situational factors, such as where the observer lives (e.g. Providence vs. Ithaca), physical factors, such as how the observer locomotes or how tall they are, but also by developmental factors, such as is the observer a child or their caregiver. Given that a niche is comprised of a number of dynamic factors that change over the lifespan,

an observer's niche is not static but changing. Building from a domain-general view, an Ecological approach emphasizes that while statistical learning might be a general, robust learning ability that is driven by statistical information that it cannot be considered in a vacuum. The importance of grounding statistical learning is especially important when considering how it can contribute to perceptual change and development.

While the interrelationship of learning and an observer's niche has been proposed broadly (e.g. E. J. Gibson, 1969; E. J. Gibson & Pick, 2000; Rovee-Collier & Cuevas, 2009), the dependence of statistical learning upon the learner's niche has not been clearly established. The current section argues that statistical learning is both grounded in an observer's niche but can potentially result in changes to their niche. In addition, this section considers how the concepts of embodiment and situated cognition contribute to the theoretical view of grounded statistical learning.

The Observer's Niche Can Affect Learning

An observer's niche can affect statistical learning in two important ways: Niches can change the statistical information present in an observer's sensory input; changes in niche will also affect what statistical information is relevant to the learner and thus what is learned.

The environment can be thought of as having an objective structure and the corresponding statistical regularities present in sensory input could be determined based on this objective structure (e.g. Christiansen, Onnis, & Hockema, 2009). However, the statistical information that is received by the learner will

be affected by their niche in this environment. Figure 1.5 provides an example of how the same environment, in fact the same events, could result in different statistical information in the sensory input as a result of developmental niche. Yu, Smith, Christiansen, and Pereira (2007) examined the visual input received by children and their caregivers in the same interaction and found large differences in the structure of sensory input that each observer received. While the caregiver's view included multiple objects in a spatial array occupying a small proportion of their visual field, the toddler's visual input tended to be dominated by the relevant objects in the interaction and sometimes was occupied by single objects in succession (Figure 1.6). This difference in sensory input is mediated in part by differences in arm length between children and adults: Children have shorter arms, so held objects occupy a greater proportion of their field of vision (Smith et al., 2011). However, there are likely a number of contributions to these differences in visual input. For example, adults hold objects closer to toddlers' faces than they would for another adult, suggesting socially-mediated differences in statistics. All of these factors (body composition, social interactions) comprising the niche of the observer can in turn affect what kind of statistical information they receive in their sensory input, and, correspondingly, what kind of structure they perceive in the world.

Differences in an observer's niche also affects what statistical information is learned even given the same sensory input. With prior experience, an observer is able to distill different perceptual information from the same sensory input. Thus, if two observers received identical sensory information, the one with prior experience would be able to pick out more information from the same sensory input. Take the example of language. Upon hearing the same utterance, the information picked out by an infant just learning the language and

Figure 1.5: Differences in developmental niche can lead to different statistical information in sensory input even for two observers viewing the same event. Pictures from Chen Yu, Indiana University.

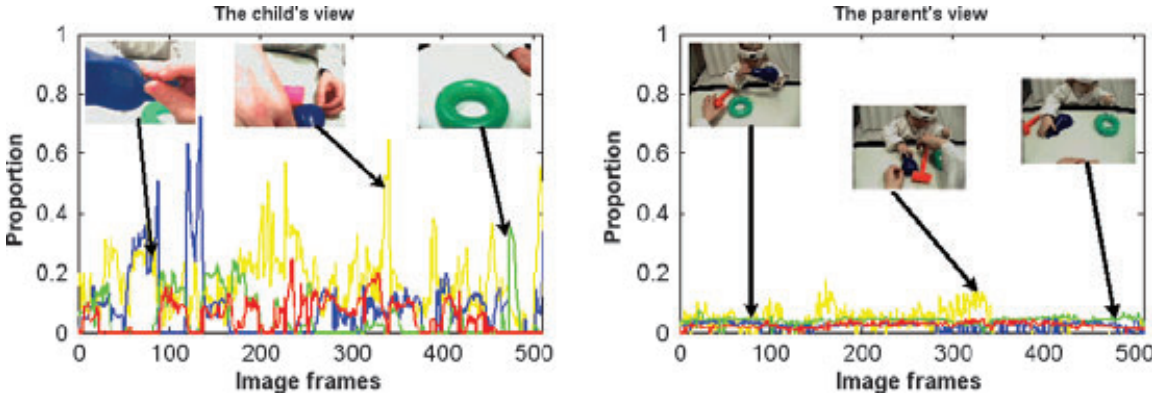


Figure 1.6: A time series of the visual input for a toddler or child (left), caregiver (right) dyad. The proportion of the field of view occupied by individual objects (green, red, blue) and body parts (yellow) is shown for each time point. From Smith et al. (2011).

an adult already proficient in the language will be different. Given that learning is a progressive and ongoing process to reduce uncertainty, the different perceptual information available to these two observers will affect what statistical information would be beneficial to learn from: The infant might determine the functional categories of the speech sounds, while the adult might focus on the relationship between the semantics of the utterance and the visual environment. In this way, the observer's niche will determine what is learned even given the

same sensory input.

While it might be possible to objectively qualify the structure, and relatedly the statistics available in sensory input, the current section argues that statistical learning tasks engaged in by an observer will be affected by their niche: Differences in niche will result in different statistical information in sensory input; moreover, even if the same sensory input is received, the current niche of the observer will affect what perceptual information is picked out of that input and thus, what patterns might be most relevant to learn in order to reduce uncertainty.

Embodiment and Situated Cognition

The concepts of embodiment and situated cognition are connected to the view that learning occurs relative to an observer's niche. *Embodiment* is essentially the view that the operations of the brain are closely tied to the operations of the body and the cycles of perception and action (e.g., Barsalou, Simmons, Barbey, & Wilson, 2003; Glenberg, 1997). Similarly, *situated cognition* emphasizes that cognition occurs within an external environment and these two systems are coupled such that they affect each other. These two terms can be considered as emphasizing the same point: that cognition cannot occur in a vacuum, that the brain must be considered in relation to the body's motor and perceptual systems and also in relation to the environment in which it is situated.

Learning Can Change the Observer's Niche

Section 1.1 argued that experience supports perceptual changes vital to tasks such as language learning and face processing. While the exact role of learning in these developmental changes is under dispute (see recent evidence for innate face processing abilities in macaque monkeys, Sugita, 2009), there is mounting evidence for the role of experience, in part mediated by statistical learning, in the development of these abilities. According to Ecological theories, such changes in perceptual abilities are part of a change in the learner's niche. For example, an observer who is able to functionally categorize the speech sounds of their ambient language exists in a different niche than another learner that is unable to do so. Thus, learning from the statistical regularities in sensory input could result in changes in the learner's niche.

In sum, this section argues that statistical learning, as it exists in the wild, must be considered in relation to the observer's niche. Specifically, what is learned from sensory input is affected by the observer's niche as the niche determines both what statistical regularities are present in sensory input, this could be considered as a form of embodied cognition. Moreover, what statistical regularities are relevant to be learned given identical sensory input are also affected by the observer's niche. In addition, statistical learning has the potential to affect the observer's niche as new information gained through statistical learning might affect future perception. However, the influence of knowledge gained through statistical learning on perception has not been directly tested in laboratory experiments and is a topic of empirical study in this dissertation.

1.5.3 The Argument and Layout of This Dissertation

The relationship between the observer's niche and statistical learning, as outlined above, could be seen to be mediated by changes in perception. The previous section argued that an observer's niche can affect what is learned based on the perception of sensory input; differences in the ability to pick out information will change what statistical regularities are available to the observer. In turn, knowledge gained through experience with statistical regularities can result in changes of the observer's niche through changing the ability of the observer to pick out information in future sensory input. In this way, the dynamic relationship between an observer's niche and statistical learning can be considered to be mediated by perceptual change. Thus, Ecological approaches suggest a role for perception in statistical learning. Similarly, for statistical learning to be considered an embodied and situated cognitive process it must be affected by sensory-motor processes and/or the environment (see Barsalou et al., 2003; Markman & Brendl, 2005).

Following from this theoretical framework, this dissertation takes a different perspective from the domain-general accounts of statistical learning presented above. The current work broadly proposes that statistical learning arises from an interrelationship between perception and learning and memory during experience with statistical regularities. Figure 1.7 presents an illustration of this framework. Specifically, during the experience with statistical regularities perception can both support and constrain what is learned and learning and memory systems can influence perception though the integration of knowledge of environmental structure gained through experience.

To support of this proposal, one must demonstrate that perceptual processes

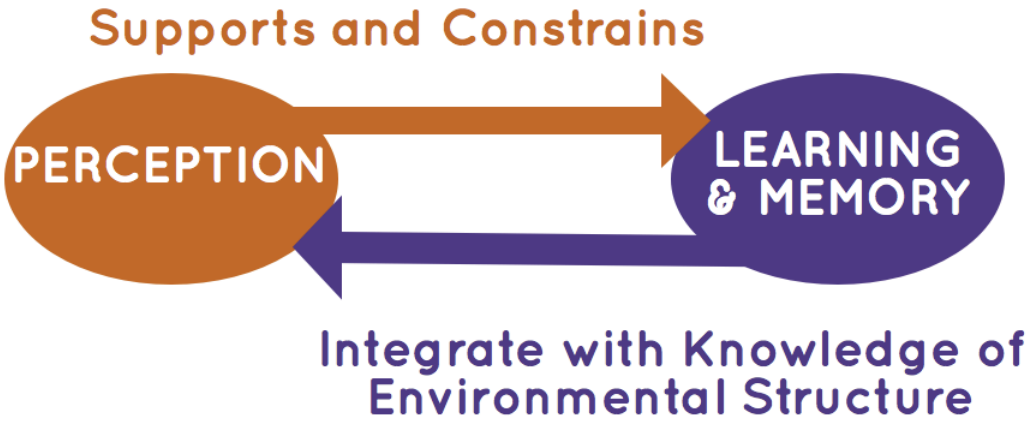


Figure 1.7: Illustration of the theoretical framework proposed in this dissertation: during experience with statistical regularities, perceptual and learning and memory processes are mutually influential.

non-trivially influence what is learned from experience statistical regularities and that learning from statistical regularities influences what is perceived. The empirical work presented in this dissertation provides evidence for both of these claims.

In Chapter 2, perceptual organization is found to both support and constrain what is learned from perceptually variable statistical patterns. This finding is discussed in relation to the application of statistical learning to language acquisition.

Chapter 3 presents a novel demonstration that perceptual identity (i.e. whether statistics are presented as auditory or visual stimuli) affects what is learned; specifically, that changes in rate of presentation have opposite effects on visual and auditory statistical learning. This chapter demonstrates that the differences in perceptual processing across perceptual modalities can affect what is learned despite receiving identical statistical information.

Chapter 4 presents evidence for a dynamic relationship between perception and statistical learning over development. Specifically, this chapter examines the relationship of auditory and visual learning in infancy and finds a different pattern than that found in adult learners (e.g. in Chapter 3). Moreover, the paper finds evidence for changes in auditory SL for infants aged eight to ten months.

Finally, Chapter 5 finds that structured experience can support changes in object perception. Neuroimaging data highlight the involvement of the medial temporal lobe, an important learning and memory system, in integrating across successive experiences to support changes in perception.

CHAPTER 2

IS STATISTICAL LEARNING CONSTRAINED BY LOWER LEVEL PERCEPTUAL ORGANIZATION?

The content of this chapter is currently under review at *Cognition*; see Emberson, Liu, and Zevin (under review)

In order for statistical information to aid in complex developmental tasks such as language acquisition, learning from higher-order statistics (e.g. across successive syllables in a speech stream to support segmentation) must be possible while organization at lower levels of description (e.g. speech categorization) is still developing. The current study is the first examination of how differences in lower level organization and higher level statistical information shape statistical learning. To this end, we presented adult participants with multiple sound exemplars from novel, complex sound categories, designed to reflect some of the spectral complexity and variability of speech. Higher-order statistical information, determined based on sound categories not individual exemplars, could support stream segmentation. However, perceptual similarity judgments and multi-dimensional scaling revealed that participants only discriminated three of four experimenter-defined categories, creating a tension between lower level perceptual organization and higher-order statistical information. We examined whether statistical learning is more consistent with statistical learning being bottom-up, constrained by the lower levels of organization, or top-down, optimal learning despite lower level organization. We consistently find that learning is constrained by perceptual organization. Moreover, participants generalize their learning to novel sounds that occupy a similar perceptual space, suggesting that learning occurs based on regions of perceptual space. Overall, these results reveal a constraint on the statistical learner where statistical information is determined based on lower level organization and thus current categorization abilities. These findings have important implications for the role of statistical learning in language acquisition.

2.1 Introduction

Starting in infancy and continuing into adulthood, humans are highly sensitive to regularities in their environment. From these regularities, which can also be conceptualized as statistical information, it is possible to learn a large amount about the structure of the world without explicit feedback or innate knowledge. In a little over a decade, learning from statistical information has been implicated in the processing and acquisition of a variety of perceptual and cognitive skills, including knowledge of causal structure and human action (Baldwin, Andersson, Saffran, & Meyer, 2008; Oakes & Cohen, 1990, 1995; Sobel, Tenenbaum, & Gopnik, 2004), visual processing (Brady & Oliva, 2008; Yuille & Kersten, 2006), and, the focus of this paper, language learning (Goldwater, Griffiths, & Johnson, 2009; Saffran et al., 1996).

It is common practice for the learning tasks and models that populate this literature to investigate learning from statistical information at a single level of description, for example, to investigate statistical learning available at a relatively low-level of the sensory signal (e.g. acquiring the internal structure of categories through distributional statistics) and in a separate study to examine learning from statistical regularity characterizing relationships at higher levels of description (e.g. segmentation of lexical items statistical information across sound categories or syllables). As a consequence, most existing statistical learning literature implicitly assumes that learning of the statistics at lower levels is resolved prior to learning from the statistics present at higher levels of description. However, developmental trajectories of statistical learning suggest that this assumption is not borne out. Furthermore, the nature of interactions between different levels of statistical learning may have important implications

on our understanding of the specific mechanisms that support learning at each level.

In the current paper, we focus on the domain of language acquisition and examine how statistical learning proceeds when lower and higher level information must be resolved *simultaneously*, a situation we argue is more akin to the task faced by the natural language learner. The *lower level* learning challenge we chose to investigate is that of acquiring auditory categories which models the problem listeners face when having to categorize substantial acoustically-variable vocal utterances into functionally-equivalent items (in their linguistic interpretation). The *higher level* learning challenge we chose to investigate is the acquisition of transitional probabilities defined across entire auditory categories that contain multiple acoustically-varying exemplars. To this end, we created a set of auditory categories, each containing acoustically-varying non-speech exemplars that are novel to adult listeners, and exposed participants to familiarization streams characterized by specific experimenter-determined transitional probabilities defined across categories. We then probed the nature of statistical learning resulting from this exposure. We contrast two broad theoretical possibilities: 1) statistical learning is primarily *top-down* – that is, statistics at higher levels can be optimally learned, regardless of how lower level information is processed; 2) statistical learning is primarily *bottom-up* – that is, the organization at lower levels is relied upon to define statistical information at higher levels regardless of whether the resulting statistics are optimal at higher levels.

2.1.1 Statistical Learning across Levels of Description in Language

Statistical information has the potential to aid language learning at many different levels of description, including speech categorization, word segmentation and lexical development, and syntactic processing (for reviews, see Romberg & Saffran, 2010; Saffran & Thiessen, 2007). The large proportion of research identified by the term statistical learning has focused on the use of transitional probabilities to accomplish word segmentation and lexical development. To illustrate, take the phrase "pretty baby" (Saffran et al., 1996) which would typically be produced as a continuous utterance /prI'tibeI'bi/. In this phrase, as well as the ambient language, transitional probabilities–the probability that one will perceive Y given the perception of X–and other statistical regularities (e.g. co-occurrence frequency) are higher between syllables that cohere to form a word (e.g. "pre" and "ty") than syllables that cross word boundaries (e.g. "ty" and "ba"). Not only can infants use transitional probabilities to segment words from a speech stream (Saffran et al., 1996; Aslin, Saffran, & Newport, 1998), but they are also more likely to use syllables linked by high transitional probability as lexical labels (Graf Estes, Evans, Alibali, & Saffran, 2007). Thus, learning from statistical regularities likely contributes to lexical development, characterized in part by the vocabulary explosion beginning around 14-months (Saffran & Thiessen, 2007).

However, the experiments that populate this literature and the corresponding models typically ignore a key challenge infants face in dealing with natural speech input: Transitional probabilities are necessarily accumulated over many different utterances of the same phrase (e.g. "pretty baby" or /prI'tibeI'bi/),

and it is well known that, even within the productions of a single speaker, there exists a large amount of acoustic variability across the multiple utterances of any given linguistic unit.[1] Essentially, each utterance "pretty baby," or any particular speech sound, has a unique acoustic waveform and is, in essence, a new exemplar of the functional category to which it belongs. Most statistical learning studies eliminate the variability across successive experiences by employing a corpus of acoustically identical repetitions of sounds (e.g. Saffran et al., 1996). Thus, this work includes the assumption that acoustic variation is already resolved, likely through the process of speech categorization – a phenomenon in which variable utterances within the same functional speech categories are treated equivalently – before transitional probabilities across the speech categories are learned.

However, reliable speech categorization is cognitively difficult and not fully developed until well beyond the age during which infants segment and learn most of their words. While there can be coherence within the distribution of acoustic properties of utterances that are functionally equivalent (e.g. within a speech category), there is also substantial overlap between the distributions of sounds that are not functionally equivalent (e.g. from different speech categories). That is, there is no clear boundary between the entire array of possible utterances intended to be one speech sound (for example, the syllable "ra") and the array of possible utterances intended to be a different, but acoustically similar, speech sound (for example, the syllable la). The distinction between all utterances of "ra" vs. "la" is defined only probabilistically, as is the case for many other speech sound contrasts. Furthermore, in the case of most speech

[1]While this paper focuses on acoustic variability, there are many other sources of information that vary across utterances, including contextual and visual information and interaction with a caregiver. All of these factors have been shown to modify cognitive processing in infancy and thus variation of these factors will alter the informational content of each utterance.

sounds, there is a lack of sufficient acoustic cues available to reliably categorize and discriminate phonemes, which have long been thought to be the basic "unit" of speech processing (Lotto, 2000; Peterson & Barney, 1952; Shankweiler, Strange, & Verbrugge, 1977). However, researchers from a broad spectrum of theoretical positions have questioned the primacy of overt, phoneme-level categorization in language use (Port, 2007; Lotto & Holt, 2000). The use of temporal windows of integration greater than the phoneme (Nusbaum & Henly, 1992; Poeppel, Idsardi, & Van Wassenhove, 2008) and dependence on communicative context (Flynn & Dowell, 1999; Obleser & Kotz, 2010) in successful speech comprehension are commonly observed. In extreme cases, people are even capable of understanding speech when the typical cues to phoneme identity are entirely unavailable (Remez, Rubin, Pisoni, & Carrell, 1981; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). Thus, the issue of how functional equivalence in speech is used productively in higher-order cognitive processes (e.g. learning from transitional probabilities across syllables) remains an unsolved problem in language research.

While it has been demonstrated that infants in the first year begin to preferentially discriminate the acoustic contrasts employed in their ambient language (e.g. Werker & Tees, 1984; Kuhl et al., 1992), it is also known that the development of speech categorization continues well beyond infancy (e.g. Hazan & Barrett, 2000; McGowan, Nittrouer, & Manning, 2004; Serniclaes, Heghe, Mousty, Carré, & Sprenger-Charolles, 2004). These results indicate that the emergence of speech categorization has a largely overlapping developmental time course with that of using transitional probabilities between syllables to segment and learn words. Even in the case where infant's show evidence of adult-like phoneme categorization and differentiation in isolation, they can fail

to differentiate these sounds during language learning tasks such as word learning (e.g. Stager & Werker, 1997; Thiessen, 2007). Thus, even early demonstrations of infants speech categorization abilities may be highly task-dependent and might not be functionally applied when learning transitional probabilities between syllables or when acquiring other language skills.

The empirical evidence suggesting fragility of early infant speech categories combined with the overlapping developmental trajectories of speech category acquisition and early lexical development cast doubt on the assumption that speech categorization is resolved before infants are learning from higher order statistical properties across speech sounds. Instead, infants must be able to learn at higher levels of description without the benefit of well-developed or supportive categorization abilities at lower levels of description. While each of these learning tasks has been extensively examined separately (i.e. the development of speech categorization and learning from transitional probabilities between syllables), learning at a higher level of description without resolution of lower level regularities has largely gone unexamined.

2.1.2 The Current Paper

We aimed to investigate learning from regularities at a higher level of description (across categories) without assuming that perceptual variability within categories has already been resolved. Specifically, we gave adult learners exposure to sound categories containing spectrotemporally complex nonspeech sounds with which the adults have no listening experience, modeling the challenge faced by infant language learners who must learn transitional probabili-

ties across speech categories while their speech categorization abilities are still developing. To this end, we examine a very simple model language where four novel sound categories are organized into experimenter-defined pairs ("words") and presented in a sound stream such that higher level transitional probabilities across units forming a "word" was always 100% and across units spanning two different words was always 50%. The novel aspect of this experiment is that each sound category is instantiated in multiple variable exemplars designed to model the natural acoustic variability that exists across any given linguistic utterance in human speech. Critically, the transitional probabilities across syllables cohering in a "word" are reliably higher as defined at the category-level, but no specific exemplar of the first unit predicts any other specific exemplar from the second unit with much higher probability than one might observe across units spanning two different "words". Thus, to reliably learn the intended category-level transitional probabilities, participants would need to resolve the within-category acoustic variability through grouping or categorizing sounds. This task would be trivial if participants somehow already had the ability to categorize or perceptually group the exemplars into the appropriate four experimenter-defined categories. However, participants' naïve perception of the sounds supports, at most, the organization of these sounds into three separate *perceptual clusters*.

Thus, this scenario creates a tension between participants' lower-level perceptual organization of sounds and the organization that they would need to have in order to learn from the intended higher-order statistical regularities. In a series of four experiments, we considered two broad theoretical possibilities for how learning might proceed:

1. Statistical learning is *top-down*. In this case, participants would be able to learn from higher order statistical regularities even if their initial lower level perceptual organization does not support it. This would result in the most robust learning of transitional probabilities defined at the higher-order category level and best possible prediction of successive sounds.
2. Statistical learning is *bottom-up*. Lower level organization is relied upon to define statistical information at higher levels. Since participants' naïve lower level organization does not support learning the intended higher-order transitional probabilities in the current experiment, participants will not have, or develop over the course of exposure, the best possible prediction of successive sounds or uncover the intended structure of the system.

2.2 Experiment 1: Naïve Perception of Novel, Spectrally-Complex Non-Speech Sounds

First, we examined in detail participants naïve perception of the sound categories used in subsequent learning studies. Specifically, listeners heard six different exemplars in each sound category. In this paper, the term *sound category* corresponds to an experimenter-defined group of acoustically varying sounds. We will use the term *sound exemplar* to refer to a single sound (out of the six in each category) heard by the participants during the experiment.

The novel sound categories were adapted from the training study of Wade and Holt (2005), which used a video game to implicitly train listeners to learn categories of novel, spectro-temporally complex non-speech stimuli. These sounds are carefully designed to capture key characteristics of natural speech

categories without sounding speech-like (to listen: `http://www.psy.cmu.edu/~lholt/php/gallery_irfbats.php`). While previous research has shown that it is possible to develop some degree of specialized processing of these sounds with adequate exposure and training (Leech, Holt, Devlin, & Dick, 2009; Liu & Holt, 2011), participants in the present study have not experienced these sounds before and thus should not, *a priori*, categorize them or process them in a specialized manner.

To characterize participants' perceptual organization of these sound categories, we asked naïve participants to perform a perceptual similarity judgment. Results from this task uncover how participants perceive the sounds in relation to each other, for example, whether they group exemplars from a single category together or intermix them with exemplars from other categories. We will refer to these perceptual tendencies broadly as a *perceptual similarity space* and as an illustration of participants' perceptual or lower level organization of the sounds and likely stems from a combination of 1) raw acoustic similarities in one or more cues characterizing the sounds, and 2) how much importance each particular cue carries for the listener. If sound exemplars are perceived as distinct from all exemplars from other categories will be referred to as *perceptual clusters*.

2.2.1 Methods

Participants

Twenty-eight students participated in the current study. All participants reported in this paper were undergraduates at Cornell University who partici-

pated in exchange for course credit. Participants were asked to report any auditory, visual, or neurological deficits via post-experimental questionnaire; no participants reported any such deficits.

Sound Stimuli

Adapted from Wade and Holt (2005), all four sound categories were designed to have two spectral peaks, P1 and P2. Each peak has a steady-state frequency portion and a transitioning frequency portion, similar to the acoustics of syllables containing a vowel and a semivowel or liquid (for a schematic diagram of the four categories of stimuli, see Figure 2.1). The two *Easy* categories (E1 and E2) each have a P1 which begins with a period of steady state frequency and then gradually decreases in frequency. Both Easy categories also have a P2 which begins with a steady state period and either consistently increases (E1) or decreases (E2) in frequency. The name *Easy* comes from the fact that these categories are theoretically discriminable from each other based on one simple acoustic cue: the direction of the frequency transition of P2. The two *Hard* categories (H1 and H2) have a P1 that begin with an increasing frequency transition followed by a steady state. They also both have a P2 that begins with a frequency transition followed by a steady state, but both H1 and H2 completely overlap in their steady-state frequencies and both contain rising and falling transitions within category. Thus, only a higher-order interaction between two cues, the onset frequency and the steady state frequency, creates a perceptual space in which H1 and H2 are discriminable from each other (see Wade & Holt, 2005, for a comprehensive discussion).

The learning curves of participants trained to categorize these sounds (Liu

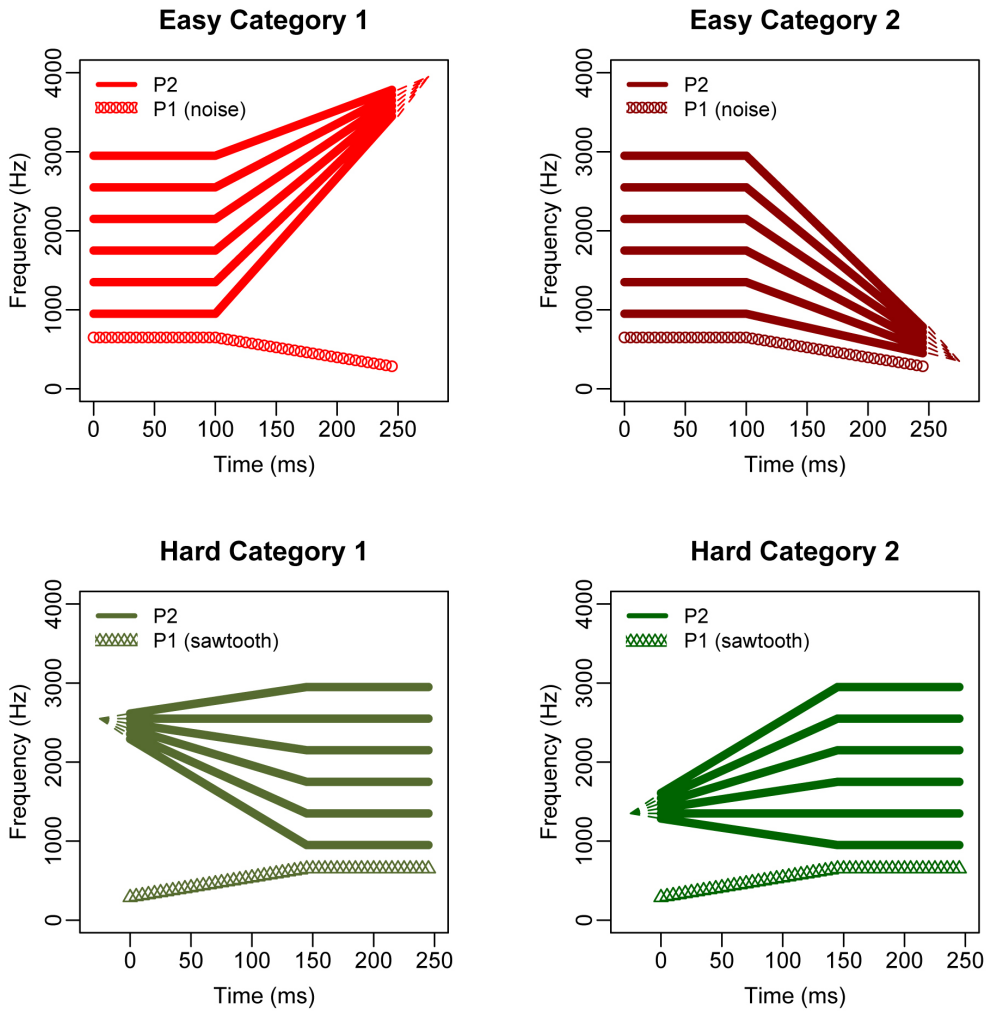


Figure 2.1: Schematic diagram of the spectrotemporal properties of the stimuli employed in all experiments. Each sound has two components: P1 (constant over all stimuli in a category) and P2 (varies for each stimulus).

& Holt, 2011; Wade & Holt, 2005) suggests it is easiest to first discriminate the Easy categories from the Hard categories and the two Easy categories from each other. Participants learn relatively quickly to categorize E1 and E2 sounds well above chance in the training paradigm, but substantially more training is necessary to produce above-chance categorization of sounds in H1 and H2. Note that it is reasonable for stimuli from Hard categories to be easily discriminable from those of Easy categories because the P2 has different carrier waves for Hard (sawtooth) vs. Easy (noise) sounds, and they have different P1 waves.

We used six different exemplar sounds from each of the four categories (e.g. from E1, they would hear $E1_1$, $E1_2$, ... , $E1_6$), all of which share a common P1 but diverge in their P2 which are illustrated in Figure 2.1.

Stimulus Presentation

All sounds were presented using over-the-ear headphones (Sony MDR-V150) at a comfortable, above-threshold volume. Instructions and stimuli were presented using PsyScope X B53 on MacMini computers. During sound presentation, participants observed blank, white screens on 17 in CRT monitors. All sounds were presented for 300 ms. Each trial began and ended with 500 ms of silence, and the two sounds were presented separated by a pause of 500 ms.

Similarity Judgment

After hearing both sounds, participants were asked to report how similar the sounds were on a scale of 1 to 4 (1 = the same and 4 = completely different) on a keyboard. Participants were given an unlimited amount of time to make their

responses.

For practical purposes, it was necessary to limit the number of trials by partitioning the full set of 24 exemplars (six from each of the four categories) into two subsets; one subset contained exemplars 1, 3, 5 from each category and the other subset contained exemplars 2, 4, 6. Half the participants performed similarity judgment on one subset, the other half on the other subset.

We performed MultiDimensional Scaling (MDS) of the aggregate data across participants using the IsoMDS routine, part of the MASS library for R (Venables & Ripley, 1999). We restricted the number of dimensions to two. ANOVAs were also run on raw similarity scores, treating subjects and items (sound exemplars) as random factors to complement the MDS results. Since participants did not repeat judgments in the perceptual similarity task and received only a subset of items to evaluate, these analyses are conducted on the responses not an aggregate of subject responses. *min F'* will be reported whenever both subject and item analyses were significant.

2.2.2 Results and Discussion

The results of the MultiDimension Scaling (MDS) of the aggregate data from all participants, presented in Figure 2.2, indicate that participants are able to perceive the Easy and Hard categories as distinct from each other: This is revealed through the linear separability of E1, E2 and H1, H2. The figure also reveals a distinction within the Easy categories: The exemplars from E1 are linearly separable from E2. This pattern of results indicates that participants also perceive a difference between the two Easy categories: E1 occupies a distinct region of

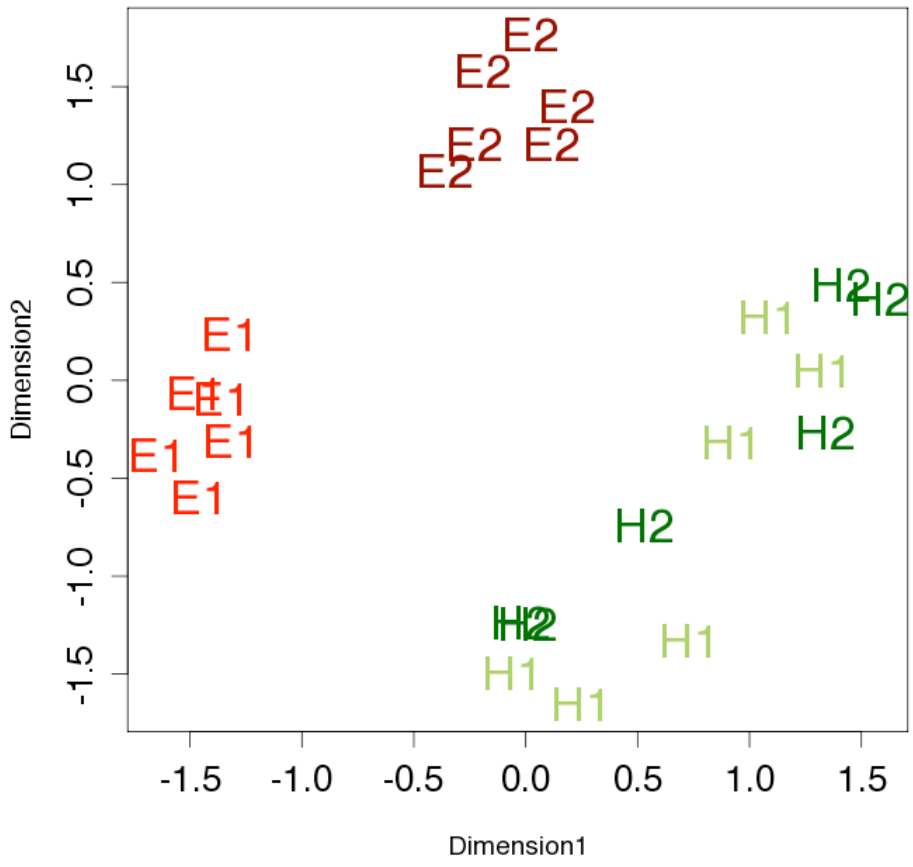


Figure 2.2: Perceptual Distance between all stimuli for all subjects. Similarity judgments were entered into MDS analysis with two dimensions.

perceptual space from E2, such that each E1 stimulus is on average more similar to all other E1 stimuli than to the E2 stimuli. This is not true for the H1 and H2 stimuli which are more or less evenly distributed across the second perceptual dimension.

We then analyzed raw similarity scores in a series of ANOVAs. First, we tested whether participants separate E from H categories by examining the resulting judgments of three possible trial types: comparison within the Easy categories (e.g. $E1_1$ vs. $E2_3$, $E1_1$ vs. $E1_5$), within the Hard categories, and a compari-

son between Easy and Hard categories (e.g. $E1_1$ vs. $H2_3$). This analysis revealed a significant effect of trial type, *min F'* (2, 234) = 22.088, $p < 0.0001$. The means for these contrasts are higher for between category types (E vs. H: 2.854) than within either E or H category trials (E vs. E: 2.243, H vs. H: 2.200) demonstrating that exemplars between Easy and Hard categories are rated as less similar than exemplars within either Easy or Hard categories.

Next, we investigated how participants perceive the four experimenter-defined categories (i.e. E1, E2, H1, H2), by crossing comparison type (Same vs. Different category) with category type (Easy vs. Hard). A significant interaction between these factors was observed, *min F'* (1, 119) = 27.955, $p < 0.0001$, such that there was a large difference in perceived similarity for Same vs. Different comparisons that was restricted to the Easy categories. That is, consistent with the MDS analysis, participants perceived the Easy items as forming two clusters roughly corresponding to the experimenter-defined categories, whereas the Hard items were equally similar between and within these categories. There also was a main effect of Same-Different judgments, *min F'* (1, 119) = 25.509, $p < 0.0001$ but no main effect of Easy-Hard trials. A planned comparison revealed that the effect of comparison type was significant only for stimuli from the Easy categories: *min F'* (1, 73) = 90.738, $p < 0.0001$.

In sum, MDS (Figure 2.2) and statistical analyses reveal that participants perceive Easy and Hard categories as distinct and they further perceive the difference between the two Easy categories (E1 and E2) but fail to perceive the difference between the two Hard categories (H1 and H2). We will refer to organization of lower level perceptual information into three categories as the *naïve* or *participant perceptual organization*. These results confirm that naïve perceptual

organization does not conform to the experimenter-defined categories.

2.3 Experiment 2: Statistical Learning across Multiple Exemplars of Novel Sound Categories

Having determined that participants' perceptual organization of the sound categories does not confirm to the four experimenter-defined categories used to create the model language, we now ask how participants' learn from statistical information defined based on these four categories of sounds. Does their perceptual organization determine what statistics are learned (bottom-up) or will participants learn based on the four experimenter-defined categories despite lacking supportive lower level organization (top-down)?

As in previous statistical learning experiments, we examined whether participants can distinguish pairs heard during familiarization from novel pairs of sounds that violate the structure of the training exposure (i.e. foils). Unlike previous experiments, but like naturalistic experience, statistical regularities were defined across sound categories with multiple exemplars (e.g. $E1_1$, $E1_2$, ... , $E1_6$).

The current learning paradigm creates an interrelationship between sound categorization and higher order statistical information by making learning based on the individual sound exemplars extremely difficult. First, there are a large number of exemplars (24 exemplars across four categories of sounds). Second, all exemplars from one sound category are presented with all others in the paired category (e.g. $E1_1$-$H2_1$, $E1_1$-$H2_2$, ...) so that a single exemplar (e.g. $E1_1$) predicts six other exemplars with equal likelihood, and a specific exemplar

pair (e.g. $E1_1$-$H2_2$) is only presented together twice during familiarization. If participants are to learn from transitional probabilities, this information must be determined based upon groups or categories of sounds not individual exemplars: While each specific exemplar pair is only presented twice (e.g. $E1_1$-$H2_2$), there are learnable transitional probabilities across these sound categories (E1-H2) as the exemplars from a single category are presented together with the groups of sounds comprising a second category many times. Thus, there is an interrelationship between grouping sounds and higher level statistics entailing 1) that how one clusters or categorizes sound at the lower level affects the resulting statistical information at higher levels and 2) that knowledge of certain higher level statistics could arrange groups of sounds at the lower level of description.

The intended structure of the stream is based on the experimenter-defined categories (E1, E2, H1, H2). Experiment 1 established that participants' perceptual organization only supports three perceptual clusters: E1 and E2 categories are separated in perceptual similarity space while the two remaining categories are not perceived as distinct creating an omnibus H cluster or category. We aim to consider statistical learning outcomes in terms of whether participants are able to learn the intended structure or whether learning is constrained by the lower level perceptual organization of the sounds.

At this point, it is important to consider the goal of the learner. Is the goal of the learner more in line with one of these theoretical outcomes than the other? We view the goal of the learner as two fold: The first goal is to uncover the intended structure. This goal is especially relevant in language learning where language structure is shared amongst members of a community, and a learner

must decipher the language so that they will be able to use these structures to communicate effectively to another. However, it is unclear how a leaner is able to determine what the intended structure is beyond what they are able to learn (e.g. the presence of statistical information). It is possible that interaction between language learners and "teachers" (e.g. caregivers) is organized in such a way as to scaffold a direction of learning (e.g. sentence frames support word recognition, Fernald & Hurtado, 2006; speaking in variation sets to support more effective word segmentation and verb learning; Onnis, Waterfall, & Edelman, 2008; Waterfall, 2006). The current experiment does not provide this kind of interactive, guided experience for the learners. They are simply exposed to the raw statistics based on a randomized stream of sound category pairs.

The second more cognitively reductionistic goal is to have the most accurate prediction of the next sound. Learning can be broadly viewed as a progressive reduction of uncertainty between internal predictions and the outcomes of events (e.g McClelland, 2002; Schultz, Dayan, & Montague, 1997). Pursuing this goal would result in statistical learning because uncovering structure allows for better prediction of successive experiences. This goal makes use of predictive processing to both reduce cognitive effort in processing correctly anticipated events and provide an error signal for incorrect predictions. This view has been applied to adult language processing (Altmann & Mirković, 2009) but such a link remains relatively uninvestigated in the domain of language learning.

Beyond the goal of learning the intended structure of the ambient language, if we characterize the goal of the learner as the best prediction, the learner should uncover the experimenter-intended structure of the model language (i.e. use the statistics based upon four sound categories). If a learner predicts the

next sound based on any six exemplars from a given sound category, the learner has a 1/6 chance of correctly guessing the next sound and is making a prediction based on a smaller region of perceptual space. If the learners prediction is based on the three categories in the naïve perceptual organization, the participant has a possible 12 exemplars to predict in the omnibus Hard category. Thus, there is less effective prediction when grouping is based on the naïve perceptual organization. An alternative possibility is that participants could predict based on the category and not the successive sounds. If participants predict successive sounds based on sound category, it would be advantageous to have larger categories thereby increasing the chance of a correct prediction. Using *reductio ad absurdum*, if there is no penalty for predicting based on larger categories, the best strategy for the learner would be to place all sounds into a single category thus always correctly predicting the next category of sound. This strategy would result in no discovery of structure or no statistical learning and would not be beneficial in reducing processing costs as the predictions are too general. Also, there is no added benefit from predicting using groupings that are smaller than the experimenter-defined categories. As summarized above, variability of exemplars within categories prevents participants from gaining predictive value from predicting based on individual exemplars (i.e. individual exemplars will only predict the next experimenter-defined category of six possible exemplars). Thus, to gain the best possible prediction of the successive utterance in the model language, participants must use the experimenter-defined categories.

In sum, this experiment examines whether participants learn based on the statistical information as defined by the four experimenter-define categories (illustrated in top row of Figure 2.3) or whether the statistical information that participants learn from is biased by their naïve perceptual organization of the

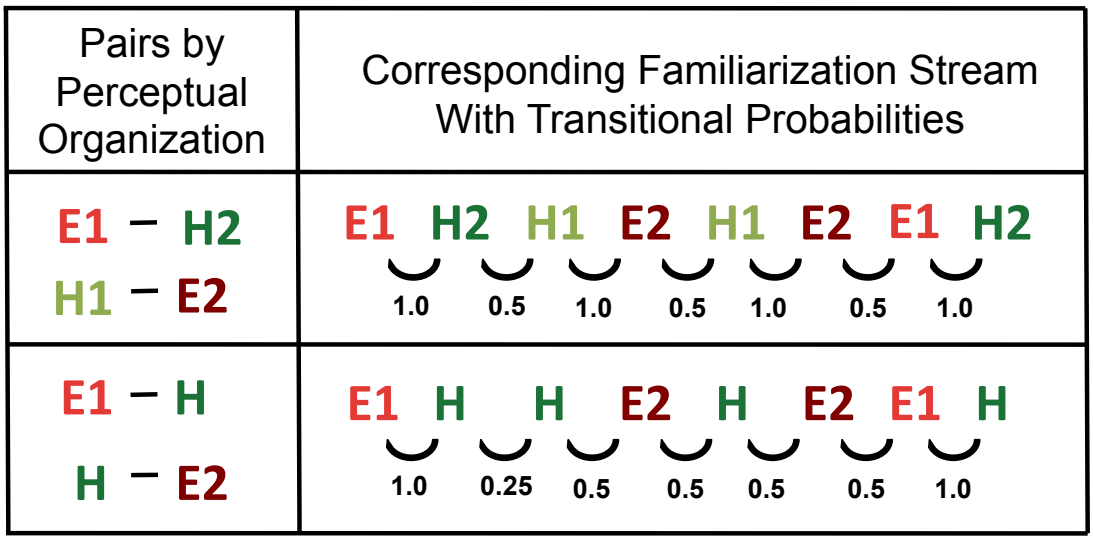


Figure 2.3: Differences in perceptual organization can change the transitional probabilities of exposure. In this experiment, perception of the four experimenter-define categories (top row) results in transitional probabilities that reflect the intended structure of the familiarization stream, but perception of only three categories (bottom row) produces transitional probabilities that do not reflect the intended organization.

sound categories as assessed in Experiment 1 (illustrated in the bottom row of Figure 2.3).

2.3.1 Methods

Participants

Fourty-five participants were recruited for this experiment. One participant was excluded for failing to complete the entire experiment.

Sound-Pair Assignment

For each participant, the four categories were counterbalanced and grouped into two pairs (e.g. participant 1: E1-H2, H1-E2, participant 2: E1-E2, H2-H1, etc.) with all possible pairings represented. We will refer to this as a sound-pair assignment. While each participant is only given one sound-pair assignment, all possible sound-pair assignments are used across participants in the experiment.

Familiarization

Sounds were presented for 300 ms each with a 115 ms inter-stimulus interval (ISI). The familiarization stream comprised 648 pairs of stimuli generated by presenting each exemplar from each category with each exemplar from its paired category twice. Pairs (in terms of experimenter-defined categories, e.g., E1-H1) were presented in a pseudorandom order so that each pair type was equally likely to be followed by both other pairs, creating the distribution of transitional probabilities shown in the top row of Figure 2.3. In addition, six instances of each sound were presented at attenuated volume (without disturbing the sequence) for a total of 144 stimuli (of 1296) that served as targets for the cover task. The duration of the familiarization period was approximately nine minutes.

Cover Task

In order to encourage participants to pay attention to the familiarization stream without explicitly asking them to track the relationships between sounds heard, a cover task was employed which consisted of participants detecting the stimuli

with attenuated volume. Participants were instructed that they would hear a stream of sounds and to press the space bar when they heard the stream get quieter. Button presses within 1.3 sec of presentation of the soft sound were considered a correct response.

Test for Statistical Learning

After familiarization, participants were given a self-timed break. After the break, they were told that they would be presented with two pairs of sounds separated by a long pause (1000 ms) and after hearing both, they would be asked to report which pair of sounds is more familiar based on their previous task. They used the *g* and *h* keys to indicate which pair was more familiar. They were also told that no new sounds are being introduced and encouraged to go with their intuition or "gut instinct". The responses were self-timed.

Participants were given 48 test trials. In each trial, one pair was composed of two exemplars consistent with those in familiarization and the other was a *foil* that violated the statistical structure from familiarization. Foils were constructed based on the four experimenter-defined sound categories: If pair 1 is AB and pair 2 is CD then the foils are CA and DB. These sound categories in the foils have a transitional probability of zero given the familiarization. Compared to studies using part-words to evaluate learning (e.g., Saffran et al., 1996), these foils should be more highly discriminable. All exemplars were heard and each pair was compared to each foil an equal number of times and counterbalanced order.

Perceptual Similarity Judgment

After completion of the SL test, participants were asked to perform a perceptual similarity judgment, as described in Experiment 1.

2.3.2 Results

Cover Task

Participants responded correctly to the "soft" sound with an average of 76% accuracy. No participants were excluded based on Cover Task performance.

Perceptual Similarity Judgments

Next, we examined how participants perceive the four experimenter-defined categories (i.e. E1, E2, H1, H2), by crossing comparison type (Same vs. Different category) with category type (Easy vs. Hard). As with Experiment 1, a significant interaction between these factors was observed, *min F'* (1, 125) = 26.322, $p < 0.0001$, such that there was a large difference in perceived similarity for Same vs. Different comparisons that was restricted to the Easy categories. That is, consistent with the MDS analysis, participants perceived the Easy items as forming two clusters roughly corresponding to the experimenter-defined categories; whereas the Hard items were equally similar between and within these categories. There also was a main effect of Same-Different judgments, *min F'* (1, 67) = 52.640, $p < 0.0001$ but no main effect of Easy-Hard trials. A planned comparison revealed that the effect of comparison type was significant only for

stimuli from the Easy categories: *min F'* (1, 73) = 90.738, p $<$ 0.0001.

Test for Statistical Learning: Overall

We begin by examining behavioral responses for evidence of learning overall. Performance was evaluated against chance (24 out of 48 or 50%). Participants were able to reliably distinguish the category pairs heard during familiarization from foils: mean performance = 27.93 (58.2%), SD = 6.65, $t(43) = 3.93$, $p < 0.001$. Thus, overall participants demonstrated sensitivity to the statistical information presented during familiarization.

Test for Statistical Learning: Generating Learning Predictions Based on Hypothetical Lower Level Organizations

Having confirmed overall learning in the current task, we went on to evaluate what lower level organization participants employed in order learn. Specifically, we aimed to differentiate two possibilities, as outlined in the introduction: One is that participants were able to learn based on the four experiment-defined sound categories and the second is that participants were constrained by their naïve lower level perceptual organization and learned based on only three categories.

As illustrated in Figure 2.3 using perceptual organization to determine the statistical information in the familiarization stream will change the statistical information a participant receives from what was intended by the experiments and what is optimal for prediction. Similarly, difference in perceptual organization can change how difficult it is for foils to distinguish from the pairs during

familarization.

Given the relationship between groupings of sounds and reliability of statistical information, we generated a set of predictions of whether a given participant would be able to distinguish pairs from foils at test. To do this, we determined the transitional probabilities for the pairs during familiarization and for the foils used at test for the three hypothetical lower level organizations. If the transitional probabilities between pairs are 1.0 during familiarization and the transitional probability of the foils averaged less than the transitional probabilities between pairs during familiarization (i.e., less than 0.5), we predicted above-chance performance in discriminating learned pairs from foils. If these conditions were not met, we predicted that participants would be unable to distinguish pairs from foils at test. A summary of predictions is presented in the table of Figure 2.4; predictions for the first experiment are presented in the left panel of the figure.

In addition to which lower level organization is used, the particular sound-pair assignment that a participant is assigned to also affects predictions for test performance. As outlined in Section 2.3.1, sound-pair assignment is how sound categories were assigned to the pairs for a single participant (e.g. participant 1: E1-H2, H2-E2, participant 2: E1-E2, H2-H1, etc.). Figure 2.3 provides an example from the E1-H1, H2-E2 sound-pair assignment. If participants use the four experimenter-defined categories to learn, participants will receive robust statistical information regardless of sound-pair assignment. This is reflected by positive predictions for learning in the first row of the table in Figure 2.4. However, there are two other ways to organize lower level information there are non-uniform grouping across four experimenter-defined categories and that in-

teract with sound-pair assignment to affect the resulting statistical information. There are three general types of sound-pair assignments where learning predictions are equal for all functional groups:

Sound-Pair Assignment 1. Participants who had both Easy categories assigned to one pair and the two Hard categories assigned to the other (i.e. EE, HH).

Sound-Pair Assignment 2. Participants who had Easy and Hard categories mixed across pairs but in consistent ordinal position in both pairs (i.e. EH, EH or HE, HE)

Sound-Pair Assignment 3. Participants who had Easy and Hard categories mixed across pairs in different ordinal position (i.e. HE EH).

Test for Statistical Learning: Determining Sound Groupings Used to Learn

Participants were organized according to three types of sound-pair assignments summarized above in order to examine differences in learning predictions. If participants learn based on the four experimenter-define categories, there should be no effect of sound-pair assignment (see first row of the table in Figure 2.4). However, a one-way ANOVA revealed a significant effect of Sound-Pair Assignment ($F(2, 41) = 7.71$, $p = 0.001$). We next tested for learning in each Sound-Pair Assignment type. This analysis revealed that Sound-Pair Assignments 1 and 2 reliably discriminated correct pairs from foils (Sound-Pair Assignment 1: mean = 28.5 (59.4%), $SD = 7.06$, $t(13) = 2.37$, $p < 0.05$; Sound-Pair Assignment 2: mean = 31.8 (66.3%), $SD = 5.83$, $t(14) = 5.18$, $p < 0.001$) whereas participants in Sound-Pair Assignment 3 failed to discriminate correct pairs from foils: mean = 23.53 (49%), $SD = 4.27$, $t(14) < 1$, *n.s.* (Figure 2.4).

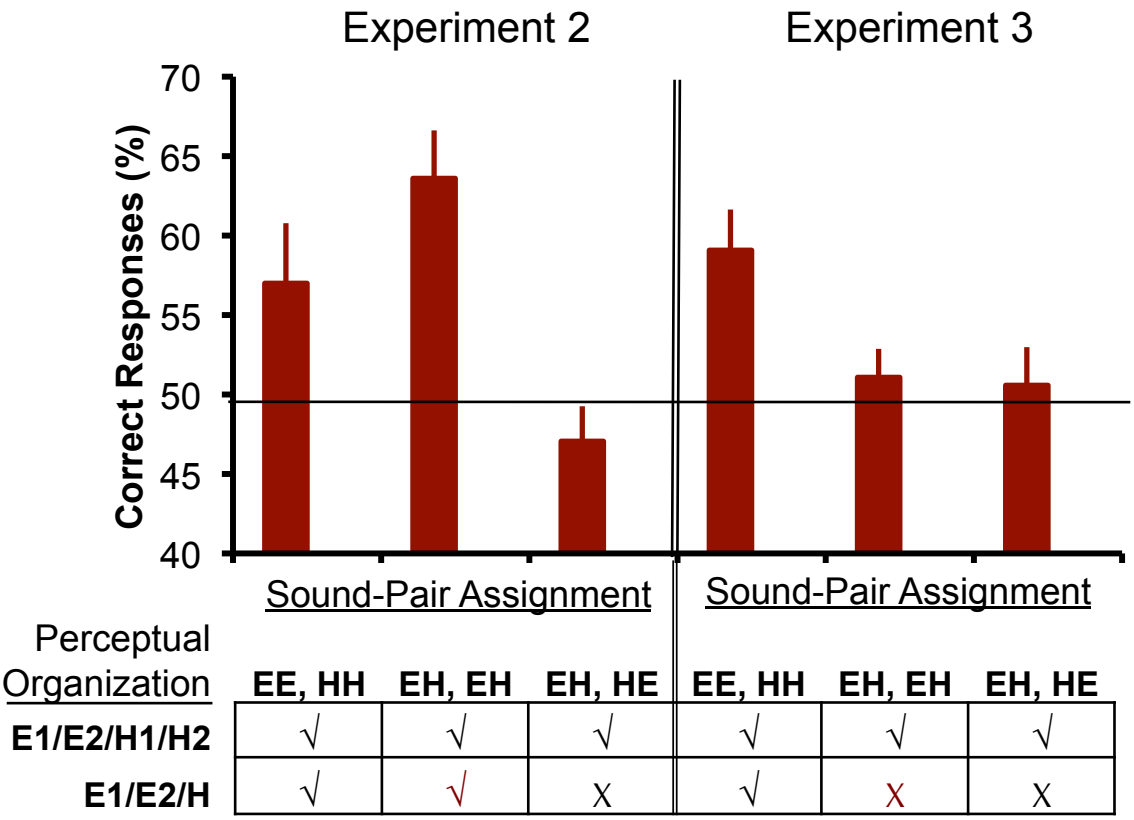


Figure 2.4: Discrimination at test for each Sound-Pair Assignment compared to predictions of whether or not we predict discrimination of pairs from foils based on both the transitional probabilities during familiarization and at test. Table: The three columns for each panel correspond to these three types of Sound-Pair Assignments, and the two rows correspond to the two possible ways to generate statistical information: top-down or consistent with experimenter-defined categories and consistent with the naïve perceptual organization. Critical predictions for the current study are presented in maroon.

Comparing this pattern of results to our predictions (table in Figure 2.4), this is evidence that participants' learning was constrained by their naïve lower level organization of the sound categories.

2.3.3 Discussion

Participants are able to learn from the temporal structure of a sound stream comprising novel, spectro-temporally complex sounds for which participants had no prior experience. This result confirms that participants are able to learn from higher level statistical information amidst perceptual variability. The model language was based on four experimenter-designed categories (E1, E2, H1, and H2). Learning based on this organization of sound exemplars would produce the best learning outcome. However, the pattern of learning observed suggests that participants' learning was constrained by their naïve perceptual organization.

2.4 Experiment 3: Generalizing the Pattern of Learning

The results of Experiment 2 were best fit by predictions based on sound-pair assignments that were initially built into the experiment as part of a counterbalancing scheme. A replication experiment was undertaken, in which different sound-pair assignments would lead to different predictions about learning. Although the familiarization stimulus streams were identical to those used in Experiment 2, different foil stimuli were used at test (see Chapter 3 and the paper in Appendix D for discussions of foils in statistical learning paradigms). The differences in foils for the current experiment produce different learning predictions. See Figure 2.5 for an illustration of the differences in foils (Type 1 for Experiment 2 and Type 2 for Experiment 3) for the critical change in learning predictions between experiments. This manipulation thus permits us to test the generality of the pattern of results observed in Experiment 2 and assess learning

Perceptual grouping assumption	Sound-pairs	Type 1 Foils	Type 2 Foils	Predicted post-test performance
4-category assumption E1/E2/H1/H2	E1 → H1 100% E2 → H2 100%	E2 → E1 0% H2 → H1 0%	E1 → H2 0% E2 → H1 0%	SP vs. Type 1 Foils: above chance SP vs. Type 2 Foils: above chance
3-category assumption E1/E2/H	E1 → H 100% E2 → H 100%	E2 → E1 0% H → H 0%	E1 → H 100% E2 → H 100%	SP vs. Type 1 Foils: above chance SP vs. Type 2 Foils: below chance

Figure 2.5: Transitional probabilities for one example sound-pair assignment (E1-H1, E2-H2). Of critical importance is that post-test performance behavioral predictions differ, when comparing Sound-pairs vs. Type 2 Foils, depending on what perceptual grouping participants are assuming in learning transitional probabilities during familiarization. This provides a clear way to use the post-test performance patterns to assess which perceptual grouping assumption participants are bringing to the learning task.

outcomes based on *a priori* predictions.

2.4.1 Methods

Forty additional participants were recruited from the same subject pool as in prior experiments. Methods were the same as in Experiment 2 with the exception of the foils used during test: If pairs are AB and CD (with A - D being the four categories of sounds), the foils in the current experiment were AD and CB (cf CA and DB in Experiment 2 or Type 1 foils in Figure 2.5; Type 2 foils in Figure 2.5). These new foils change the calculation of whether participants in a given sound-pair assignment and set of functional units will be able to distinguish pairs from foils at test (i.e. prediction of learning). Based on these new foils, *a priori* predictions of discrimination at test were generated in the same

fashion as Experiment 2 and are reported in the right panel of the table in Figure 2.4.

2.4.2 Results and Discussion

Cover Task Results

Participants responded correctly to the "soft" sound with an average of 73% accuracy. No participants were excluded based on their cover task performance.

Perceptual Similarity Results

As in Experiments 1 and 2, a significant effect on perceptual similarity judgments was found in analyses considering the Easy and Hard stimuli, *min F'* (2, 189) = 41.421, $p < 0.0001$. The means for these contrasts are higher between category types (E vs. H: 3.000) than within either E or H category trials (E vs. E: 2.288, H vs. H: 2.040). These results indicate that participants make a distinction between Easy and Hard exemplars.

Next, we investigated whether participants make further distinctions by examining the effects of same-different judgments within Easy and Hard categories. Following from the previous results, there was a main effect of Same-Different judgments, *min F'* (1, 135) = 18.223, $p < 0.0001$ but no main effect of Easy-Hard trials. In addition, there was an interaction of Easy and Hard with Same-Different judgments: *min F'* (1, 133) = 21.910, $p < 0.0001$. Examining Same vs. Different judgments separately for Easy and Hard categories, there was no

effect of Same vs. Different judgments for the Hard category trials, but a significant effect of Same vs. Different trials in the Easy categories: *min F'* (1, 82) = 42.169, $p < 0.0001$. These results indicate that participants perceive the difference between the two Easy categories but not the difference between the two Hard categories, as in both previous experiments.

Test for Statistical Learning: Results

As in Experiment 2, participants correctly discriminated pairs from foils overall (mean = 26.4, SD = 4.63, $t(39) = 3.31$, $p < 0.01$). A one-way ANOVA revealed that performance was not uniform across Sound-Pair Assignment ($F(2, 39) = 5.46$, $p < 0.01$). Next, we examined learning for each Sound-Pair Assignment to evaluate the pattern of test performance against the a priori predictions for three possible lower level organizations of the sounds: Learning was found for Sound-Pair Assignment 1, mean = 29.54, SD = 4.63, $t(12) = 4.31$, $p = 0.01$, but not for the other assignments, both $ts < 1$ (Figure 2.4). This pattern of results is consistent only with the predictions of learning based on the Naïve Perceptual Organization (E1, E2, and H, the second row of the table in Figure 2.4). Thus, the findings of the current experiment confirm that, while participants use the full extent of their perceptual organization captured by the perceptual similarity judgments, their learning is constrained by the lower level organization brought into the task.

2.5 Experiment 4: Generalization to New Sounds

Experiments 2 and 3 demonstrated that participants' learning of higher-order statistical learning is constrained by their lower level perceptual organization of these sounds. However, it is unclear exactly how participants are learning given that participants have no prior experience with these sounds and no specialized or explicit categorization abilities. However, we do find that participants have a non-uniform organization of perceptual space that supports the grouping of exemplars for some categories and fails to differentiate others. We find that this organization constrains statistical learning, but do participants learn from higher order statistical information defined based upon these regions of perceptual space? In the current experiment, we examine a crucial prediction of this view: If learning is indeed based on naïve perceptual organization, participants should be able to generalize knowledge gained through statistical exposure to novel exemplars that occupy a same region of perceptual space. To test this hypothesis, we familiarized a new group of participants using identical methods to Experiment 2. At test, however, both the target and foil stimuli were generated using exemplars from the four experimenter-defined categories that were never presented during familiarization.

2.5.1 Methods

Thirty-seven additional participants were recruited for this experiment. We used the same experimental procedures as Experiment 2 except for one crucial difference: Novel exemplars from the four experimenter-defined categories were used to assess learning during test. The experimenter-defined sound cate-

gories each have 11 possible exemplars that vary incrementally based on the P2 component (Wade & Holt, 2005, see Figure 2.1). We used the odd-numbered exemplars during familiarization in all previous experiments. During the current learning test, we employed the even-numbered exemplars to test for generalization of knowledge to novel sounds.

To verify that these novel exemplars are in fact grouped in the same manner as the familiarization exemplars, participants performed the same perceptual similarity test previous experiments with the novel exemplars. In order to reduce the length of the perceptual similarity task, each participant receives a subset of exemplars for each category to compare with all other selected exemplars. Because there is one fewer novel exemplar, the middle exemplar (6) was used in both subsets (either 2, 6, 8 or 4, 6, 10).

The statistical learning test was identical to Experiment 2, with the exceptions of using the novel exemplars and a reduction of total test trials due to the smaller number of novel exemplars (40 test trials). Thus, if participants are able to generalize their knowledge to novel exemplars which occupy the same perceptual similarity space, we should observe the same pattern of results as in Experiment 2: We predict learning in Sound-Pair Assignments 1 and 2 but not in Sound-Pair Assignment 3.

2.5.2 Results

Cover Task

Participants performed the cover task with an average of 78% accuracy. No participants were excluded based on cover task performance.

Perceptual Similarity of Novel Exemplars

Participants performed the perceptual similarity judgments after completing the learning test using the sounds used to test generalization (i.e. sound used during test but not heard during familiarization). Results from the perceptual similarity judgments are consistent with our assumption that the novel exemplars are grouped in the same way as the exemplars tested in Experiments 1-3. A one-way ANOVA examining the effects of three types of perceptual similarity trials (within Easy or Hard categories: E vs. E, H vs. H, and between Easy and Hard categories: E vs. H) revealed a significant effect of contrast type, *min F'* (2, 181) = 35.314, $p < 0.0001$. The means for perceptual similarity judgments are higher for between category types (E vs. H: 2.901) than within either E or H category trials (E vs. E: 2.051, H vs. H: 2.092). We then examined how participants perceive the four experimenter-defined categories (i.e. E1, E2, H1, H2), by crossing comparison type (Same vs. Different category) with category type (Easy vs. Hard). As with previous experiments, a significant interaction between these factors was observed, *min F'* (1, 101) = 19.486, $p < 0.0001$, such that there was a large difference in perceived similarity for Same vs. Different comparisons that were restricted to the Easy categories. There also was a main effect of Same-Different judgments, *min F'* (1, 71) = 41.327, $p < 0.0001$ but no main effect of

Easy-Hard trials.

Test for Statistical Learning

Test performance is consistent with the hypothesis that participants are able to generalize knowledge to novel sound exemplars. Overall, participants demonstrate significant learning, mean = 22.4 out of 40 (56%), $t(36) = 2.044$, $p = 0.048$. The predictions for learning according to Sound-Pair Assignment and level of perceptual grouping are identical to Experiment 2 (see left panel of Figure 2.4). Broken down by Sound-Pair assignment, significant learning was observed for in Sound-Pair Assignment 1, mean = 24.08 (60.2%), $t(11) = 2.512$, $p = 0.029$ and in Sound-Pair Assignment 2, mean = 25.31 (63.3%), $t(12) = 2.284$, $p = 0.041$. There is a marginally significant but below chance-level performance for the Sound-Pair Assignment 3 group, mean = 17.5 (43.8%), $t(11) = -2.147$, $p = 0.055$.

2.5.3 Discussion

In sum, the current experiment examines whether participants can generalize knowledge gained from statistical learning to novel exemplars occupying the same perceptual space. To do so, we examined whether we get the same pattern of learning as Experiment 2 when the statistical learning test is administered using novel exemplars. The current experiment essentially replicates Experiment 2 which demonstrates that statistical learning during familiarization generalizes to stimuli drawn from the same regions of perceptual space.

We hypothesized that participants use the regions of perceptual space in

their naïve perceptual organization to learn statistical regularities defined at a higher level of description. The ability to generalize knowledge to new sounds occupying the same perceptual region or cluster is a crucial prediction of this hypothesis. Thus, we find further evidence to support the view that statistical learning in the current task proceeds based upon lower level perceptual organization.

2.6 General Discussion

The current study examined a scenario where lower level perceptual organization does not readily support the learning of intended higher order statistics. This captures an important feature of the task faced by the infant language learner: Infants learn the words of their native language, based in part on statistical information defined over sequences of speech sounds, before their ability to categorize speech sounds is fully developed. Our goal was to examine whether statistical learning is constrained by lower level organization (bottom-up) or if optimum learning of the intended structure will proceed despite the current lower level organization (top-down). We consistently found evidence that participants' learning was constrained by their lower level perceptual organization, indicating that statistical learning is more bottom-up than top-down. This finding suggests that learning from higher-order statistical regularities does not readily proceed if it is not readily supported by lower level perceptual organization. This finding has important implications for the role of statistical learning in the development of complex cognitive tasks such as language acquisition. These implications will be discussed in detail in this final section.

One key assumption of this work is that adult learning from novel, complex sound categories is at some level analogous to learning from speech in the first year of life. Broadly, this assumption is based on the view that speech sound categorization is a form of perceptual expertise. It follows that the speech categorization abilities possessed by adults do not effectively transfer to novel, non-speech sound categories. Thus, the cognitive processes of an adult listening to novel sound categories are matched to an infant listening to speech sounds. While this assumption is shared amongst many studies in the statistical and language learning literature, it is important for future research to establish these same effects using speech sounds in infant language learners.

To our knowledge, this is the first investigation of statistical learning where participants must resolve information across multiple levels of description. While two recent studies employed variable exemplars of visual stimuli, these stimuli were drawn from categories with which participants have had considerable experience: Brady and Oliva (2008) used visual scenes organized by well-learned semantic categories (e.g. kitchens and beaches), while Baldwin et al. (2008) used pictures common actions such as pointing and grasping. In each of these cases, it is very likely that participants organization of these exemplars would confirm the experimenter-defined categories before the experiment providing perceptual support for the experimenter-defined, higher-order statistics. In the auditory domain, a small number of statistical learning studies have used variable natural productions of speech sounds (e.g. Thiessen, Hill, & Saffran, 2005), but neither the acoustics nor the perceptual similarity of the stimuli have been systematically controlled or examined. Thus, it is unclear how these variable productions were perceived by the learner, and what effect they had on the learning task. In contrast, the current study confirmed participants' per-

ceptual or lower level organization of the multiple exemplars and found that they did not conform to the experimenter-defined categories used to produce higher-order statistical information.

2.6.1 Implications for Language Acquisition

The current learning paradigm is motivated by the following developmental problem: How are infants able to learn their first words without the use of established, adult speech categorization abilities? The developmental trajectories of word learning and speech categorization substantially overlap. While the possibility of simultaneous development of speech sound categorization and word learning has been acknowledged (e.g. Swingley, 2008; Saffran & Thiessen, 2007), the canonical account of how statistical learning contributes to language acquisition often assumes a sequential development from lower to higher levels of description in the speech signal. Moving forward, it is essential to resolve how learning proceeds when information must be resolved across levels of description, and specifically when there is tension between lower and higher level statistical information.

The current study finds that learning can still occur despite conflict between lower level organization and higher level statistical information. Specifically, participants can gain some knowledge of the structure of the familiarization stream by relying on their naïve lower level perceptual organization to define higher-order statistics. First, it is important to note that participants do not have specialized processing for the current sounds. Thus, we find that learning can proceed over a variable acoustic stream without specialized categorization

abilities. Instead, we find that higher-order statistical knowledge, in the current task, is possibly based on clusters or regions in their perceptual similarity space. Specifically, Experiment 4 presents evidence that participants can generalize knowledge gained from statistical learning to new sounds that occupy the same perceptual similarity space.

If statistical learning can occur based on graded regions of perceptual space, this has important implications for language learning. Instead of being based on explicit or specialized categorization, perceptual clustering is likely based on participants' orientation towards acoustically salient features. Recent work has found acoustic contrasts with large, perceptually separable differences are easily distinguished by infants across the first year of life, whether or not these acoustic differences are part of their native environment (Narayan, Werker, & Beddor, 2010). While these results suggest that infants might be able to use salient acoustic differences to distinguish functional categories of sounds, it has yet to be determined whether infants can use relative perceptual similarity to group variable exemplars of sound categories and to subsequently apply this in an online task of tracking statistical regularity. However, current results suggest that it is possible that perceptual salience, by supporting lower level organization, provides an initial entrée into learning the structure of the speech stream.

While we do find evidence for learning in the current task, the pattern of learning is unconstrained by perceptual organization. If participants were able to learn using the higher-order statistics based upon the experimenter-defined categories (top down), they would have been able to a) demonstrate learning uniformly across all sound-pair assignments, b) uncover the intended structure of the acoustic stream and c) achieve the best possible prediction for successive

sounds. Instead, we find that statistical learning of higher-order regularities is non-uniform across sound-pair assignments. In fact, the pattern of learning across assignments is consistent with the three categories represented in participants' lower level perceptual organization. This finding entails that statistical learning is constrained by perceptual organization and thus is more bottom-up than top-down. In other words, the intended and optimal learning of higher-order statistics is restricted by lower level organization.

Considering this result in relation to language acquisition, one of the immediate implications us that the statistical information present in the speech stream must be considered in relation to the infants' current sound categorization abilities. Previous modeling work has attempted to objectively determine whether the statistical information available to the infant (e.g. the CHILDES corpus, MacWhinney, 1991) is sufficient for certain aspects of language learning (e.g. Christiansen et al., 2009). However, these models rely upon the processing of the speech signal according to categorization abilities that the infant does not possess. Despite positive findings in these studies, it is likely that this statistical information is not immediately available to language learners because of their current perceptual organization. Thus, an infant's perceptual organization and categorization abilities must be considered when considering what abilities statistical information might support during language acquisition.

If the uncovering of language structure is indeed constrained by the organization of speech sounds, the early stages of language development might be more dynamic than has been previously characterized. We find evidence that statistical information does affect learning tasks across levels of description with lower levels constraining higher levels. Thus, at any given moment, an

infant's knowledge of the structure of language is filtered based on her organization of speech sounds. However, we know that lower level organization has a protracted development with changes occurring well beyond infancy (see Section 2.1). Current findings suggest that each change in speech categorization abilities could potentially have a cascading impact on the statistical information available to the learner. This view presents the statistical language learner as having a much more re-iterative and dynamic development than a sequential view of language learning from lower to higher level statistical information.

In this dynamic view of language learning, learning based upon perceptually salient differences could drive changes of functionally-equivalent contrasts that are not supported by initial perceptual organization. A recent study with infants suggests such a possibility: Thiessen (2007) demonstrated that phonemic contrasts that are indistinguishable on their own in a word learning task (daw/taw) are learnable when they are previously associated with more salient phonemic distinctions (dawbow/tawgoo). According to Thiessen, the acoustically salient difference between "bow" and "goo" provides a differential context which can in turn support cognitive separation of the minimal pair of "taw/daw" necessary for subsequent word learning. These results provide some initial evidence that learning based on perceptually distinct sounds can support the acquisition of more difficult acoustic contrasts.

According to Thiessen (2007), this finding is consistent with a distributional account of the development of speech perception. According to this view, distributional statistics, based on lower level occurrence of the sound tokens, can shape the perceptual representation of sound categories (see Clayards et al., 2008; Maye et al., 2002; Yoshida et al., 2010). There is another possible mech-

anism that could support this finding: It could be that functional differentiation of sounds is supported by higher-order statistical information in the language, such as the type of co-occurrence statistics manipulated here (e.g., Beckman & Edwards, 2000). Feldman, Griffiths, and Morgan (2009) present evidence from a Bayesian model that lexical acquisition can aid in distinguishing the highly overlapping speech categories when the categories are lexically contrastive. In other words, the higher-order statistical information (i.e. reflective of lexical structure) can be used to modify representation and processing of lower-order speech categories. This model presents one exciting alternative to sequential learning across levels of description in a perceptually and informationally complex signal, like language and is potentially complimentary to a distributional account of the development of speech categorization.

In the current study, higher-order transitional probabilities functionally distinguish category distinctions now represented in participants naïve perceptual organization (i.e., H1 and H2). While we did not find evidence that exposure facilitated perceptual distinction between H1 and H2, this is not surprising given the difficulty of learning this contrast and the length of training required to shift behavior and perceptual treatment of these sounds in previous work (e.g., Leech et al., 2009; Liu & Holt, 2011). Thus, this lack of perceptual shift could be due to 1) our relatively short exposure to the higher-order transitional probabilities; 2) the fact that the presentation of within-category sounds had a flat distribution because natural speech categories exhibit a more distinctive within-category distributional structure which may make it easier to acquire categories in general; or 3) the possibility that higher-order transitional probabilities simply do not facilitate sound category learning. While previous studies and models reviewed above suggest that the last possibility is unlikely, further research

is needed. In general, our results suggest that learning functional equivalence classes based on higher-order transitional probabilities is at least harder than learning transitional probabilities derived from perceptual similarity spaces.

Models of language learning have relied on resolving the speech signal into discrete units as a method to resolve perceptual variability inherent in the speech signal and suggested that statistical language learning is a sequential acquisition from lower- to higher-order statistical information. These approaches have 1) lead the field to underestimate the potential interaction of learning at multiple levels of description in the speech signal and 2) to largely ignore the role of perceptual organization and perceptual salience of contrastive features in language learning. The current results demonstrate that the issue of perceptual salience and the interaction of statistical regularities defined at different levels of description is important to consider in the investigation of language learning.

2.6.2 Learning from Statistical Regularities in a Perceptually Variable World

A central problem to establishing how experience supports the development of complex behavior is understanding how the learner is able to assimilate information embedded in the highly variable and complex natural environment. Similarity, the development of effective knowledge in a variable world requires the ability to generalize. In the current experiments, we demonstrate that adults cannot only learn from higher-order transitional probabilities based on perceptually variable exemplars, but they are also able to generalize this knowledge to other stimuli that occupy the same region of perceptual space.

It is possible that the acquisition of generalizable knowledge is directly related to having received perceptually variable experience. Recent studies examining infant learning present some corroborative evidence for generalization based on variable experience. Lany and Saffran (2010) exposed infants to novel labels that have a similar morphological structure (e.g. three syllables) and then created an association between a subset of these labels and a number of visual referents from a single object category. They found evidence that infants at 22-months-old were able to generalize this information to unheard exemplars sharing these morphological characteristics. While not explicitly tested, variable experience during exposure may have facilitated generalization of these higher-order statistical regularities in infants. Relatedly, rule-learning studies present infants with a higher-order sequence where each new instance is perceptually novel and find that infants can generalize this sequence to entirely novel sounds or pictures (e.g. Marcus et al., 1999; Saffran et al., 2007). These rule-learning studies are at an extreme level of both perceptual variability during exposure and ability to generalize knowledge. In the current experiment, learners received exposure to variable information throughout learning, and possibility because of the variability of experience, they were able to generalization based of perceptual organization.

While the relationship between variability and generalization remains to be directly established in relation to learning from environmental statistics or regularities, this relationship has been established in related domains (e.g. E. J. Gibson, 1969). Recent work in word learning has explicitly examined the role of acoustic variability: These results suggest that hearing labels produced by multiple speakers aids in learning label-object pairings (Rost & McMurray, 2009). Moreover, Clopper and Pisoni (2004) linked speaker variability and successful

speech perception. Overall, this work suggests that a learner may be better able to extract abstract or generalizable relationships between stimuli with increased perceptual variability during exposure.

While it seems intuitive that perceptual variability would obfuscate the statistical patterns, it may be that the variable experience that learners experience in their everyday environment may support more robust, behaviorally-effective learning within a complex environment. Specifically, variability of experience could allow infants to distinguish which aspects of the sensory signal are an integral part of a higher-order pattern or invariant (E. J. Gibson, 1969) without having to *a priori* attend to the part of the relevant incoming information such as the sequence of syllables (Saffran et al., 1996) or an abstract rule supporting the choice of dog pictures (Saffran et al., 2007). In fact, without variability of experience, it is unclear how infants are able to focus on some aspects of the signal and uncover statistical patterns. Such an ability, if developed, could support changes in statistical learning over developmental time and differences in learning across types of input.

An alternative possibility is that, regardless of perceptually variable experience, learners will be able generalize knowledge gained through experience with statistical regularities to stimuli of the same perceptual space. This would result in a *generalization gradient* for this statistically-based knowledge. In learning theory, as a stimulus becomes more perceptually distinct from the conditioned stimulus, conditioned responses fall off according to a generalization gradient (e.g., Guttman & Kalish, 1956). It is possible that an analogous effect exists for knowledge gained through exposure to higher-order statistical information: As sounds decrease in perceptual similarity, generalization of learning

will decrease in a related fashion, and once a certain boundary of perceptual space is exceeded, participants will fail to generalize altogether. Future research is needed to more directly address the relationship between perceptual similarity, variability of experience, and generalization of knowledge in learning from statistical regularities.

2.6.3 Conclusions

In order for statistical information to aid in complex developmental tasks such as language acquisition, learning from higher-order statistics must be possible while organization at lower levels of description is still developing. The current study is the first attempt to examine how such learning would proceed. Broadly, our results suggest some constraints on the statistical learner in this scenario: If higher-order structure is not supported by lower level organization, then the structure is not readily uncovered. Thus, statistical learning appears to be more bottom-up than top-down (i.e., what is learned at higher levels is dependent upon the organization at lower levels). This finding has important implications for the conceptualization of how statistical learning contributes to language acquisition. First, statistical information must be considered based on the current categorization abilities of the learner not according to adult categorization abilities because the differences in categorization abilities might obfuscate statistical information and thus may not be available to the infant. This entails that the statistical information about the ambient language could be radically different for an infant and an adult. Second, the interrelationship of lower level perceptual organization and statistical information present at higher levels suggest a more dynamic and re-iterative process of language learning than previously

conceived. Specifically, each change in lower level organization can potentially reveal new statistical information at higher levels of description. Finally, the current results suggest an important role for perception, and specifically perceptual salience as an early and lasting influence on language learning.

2.7 Acknowledgements

Thanks to Dani Rubinstein, David Kalkstein, Noa Hertzfeld, and Andrew Webb for data collection. Thanks to Drs. Lori Holt and Travis Wade for use of their stimuli and special thanks to Dr. Michael Spivey. This work was funded by in part by R01 DC007694 and R21 DC008969.

CHAPTER 3

TIMING IS EVERYTHING: CHANGES IN PRESENTATION RATE HAVE OPPOSITE EFFECTS ON AUDITORY AND VISUAL IMPLICIT STATISTICAL LEARNING

The content of this chapter is published in the *Quarterly Journal of Experimental Psychology*; see Emberson, Conway, and Christiansen (2011)

Implicit statistical learning (ISL) is exclusive to neither a particular sensory modality nor a single domain of processing. Even so, differences in perceptual processing may substantially affect learning across modalities. In three experiments, statistically equivalent auditory and visual familiarizations were presented under different timing conditions that either facilitated or disrupted temporal processing (fast or slow presentation rates). We find an interaction of rate and modality of presentation: At fast rates, auditory ISL was superior to visual. However, at slow presentation rates, the opposite pattern of results was found: Visual ISL was superior to auditory. Thus, we find that changes to presentation rate differentially affect ISL across sensory modalities. Additional experiments confirmed that this modality-specific effect was not due to cross-modal interference or attentional manipulations. These findings suggest that ISL is rooted in modality-specific, perceptually based processes.

3.1 Introduction

Implicit statistical learning (ISL) is a phenomenon where infant and adult behavior is affected by complex environmental regularities seemingly independent of conscious knowledge of the patterns or intention to learn (Perruchet & Pacton, 2006). Because young infants are sensitive to statistical regularities, ISL has been argued to play an important role in the development of key skills such as visual object processing (Kirkham et al., 2002) and language learning (Saffran et al., 1996; Smith & Yu, 2008). Underscoring its importance for development and skill acquisition, ISL has been observed using a wide range of stimuli from different sensory modalities and domains (nonlinguistic auditory stimuli: Saffran, 2002; Saffran et al., 1999; tactile stimuli: Conway & Christiansen, 2005; simple visual stimuli: Fiser & Aslin, 2001; Kirkham et al., 2002). Together, these findings indicate that ISL is a domain-general learning ability spanning sense modality and developmental time.

Given that ISL occurs with perceptually diverse input, many influential models and theories of ISL have presupposed a mechanism that treats all types of input stimuli (e.g., tones, shapes, syllables) as equivalent beyond the statistical structure of the input itself (e.g., Altmann et al., 1995; Perruchet & Pacton, 2006; Reber, 1989; Shanks et al., 1997). While great strides have been made under this equivalence assumption, there is evidence, contrary to this view, that ISL is not neutral to input modality. Instead, the perceptual nature of the patterns appears to selectively modulate ISL.

In this paper, we employ a known perceptual phenomenon to examine ISL under different perceptual conditions. Specifically, we manipulated the tempo-

ral distance of successive stimuli in auditory and visual ISL streams. The perceptual literature predicts that changes of temporal distance will have opposite effects on auditory and visual processing. If ISL were also differentially affected by temporal distance, this would suggest that the mechanisms mediating ISL do not in fact treat all types of perceptual input equivalently.

In addition, we investigated the role of selective attention in modifying learning under these different perceptual conditions. While previous research has suggested that selective attention can compensate for perceptual effects in ISL (e.g., Baker et al., 2004; Pacton & Perruchet, 2008), this claim has only been tested in a small range of perceptual conditions in the visual modality only. Here we examine whether selective attention can compensate for large differences in rate of presentation in both the visual and the auditory modalities. Specifically, we predict that while selective attention may be able to support learning amidst mild disruptions to perceptual processing (as in Baker et al., 2004), attention is not sufficient to overcome more substantial changes in perceptual conditions like those explored in the current study.

In sum, we manipulated attention to auditory and visual streams under temporally proximal and distal conditions in order to examine what effect changes of presentation rates have on auditory and visual ISL. If the mechanisms of ISL are sensitive to the perceptual nature of stimulus input beyond statistical structure, then we predict that rate and modality will interact to affect learning outcomes.

3.1.1 Modality effects in implicit statistical learning

While ISL is perceptually ubiquitous, with adults and infants able to detect statistical regularities in multiple sensory modalities, recent studies with adult learners have pointed to systematic differences in ISL across these modalities (Conway & Christiansen, 2005, 2006, 2009; Robinson & Sloutsky, 2007; Saffran, 2002). Specifically, modality differences in ISL appear to follow the visual:spatial::auditory:temporal characterization seen in other perceptual and cognitive tasks, where spatial and temporal relations are processed preferentially by the senses of vision and audition, respectively (B. Kubovy, 1988).

While temporal and spatial information are both important for visual and auditory processing, these sources of information appear to play different roles across perceptual systems. The visual:spatial::auditory:temporal analogy (B. Kubovy, 1988), used to explain auditory and visual processing differences, has its roots in the nature of sensory objects. Sound is a temporally variable signal, and, since sounds do not persist, their locations in space are ephemeral. Conversely, visual objects are more spatially constant. Thus, it is adaptive for auditory processing to be more sensitive to the temporal aspects of environmental information (Chen, Repp, & Patel, 2002) whereas the adult visual system appears to preferentially encode spatial information (Mahar, Mackenzie, & McNicol, 1994). Furthermore, the visual:spatial::auditory:temporal characterization extends beyond perceptual tasks to memory (serial recall: Penney, 1989).

[1]

[1]The range of visual processing explored in the current paper is restricted: We are examining visual processing and learning of sequentially presented, unfamiliar abstract shapes. Other visual tasks have revealed the visual system to have sophisticated temporal processing (e.g., rapid serial visual presentation of scenes and photographs in Potter, 1976). However, with the current visual task, it is well established that visual processing is relatively poor especially when compared to auditory processing.

These differences in processing between auditory and visual systems are also present in ISL. Consistent with a spatial bias in visual processing, visual learning is facilitated when stimuli are arrayed spatially (Conway & Christiansen, 2009; Saffran, 2002). When stimuli are presented in a temporal stream, auditory learning is superior to vision (Conway & Christiansen, 2005). These findings point to important differences in the ways in which auditory and visual statistical patterns are learned.

We propose that comparisons of learning across perceptual modalities help elucidate the nature of the mechanism(s) underlying ISL. Moreover, these modality effects in ISL may indicate that the underlying mechanisms are sensitive to the perceptual nature of the input beyond statistical structure. One could think of these mechanisms as being "embodied" (Barsalou et al., 2003; Conway & Christiansen, 2005; Glenberg, 1997) where the learning mechanisms are situated in the perceptual process itself.

3.1.2 Modality-specific perceptual grouping and ISL

Modality differences can also be conceptualized through the lens of Gestalt perceptual-grouping principles. The spatial bias in visual processing has been formalized by the "law of proximity": Visual stimuli occurring close together in space are perceptually grouped together as a single unit (M. Kubovy, Holcombe, & Wagemans, 1998; Wertheimer, 1938), with the strongest grouping occurring in spatially contiguous visual objects (Palmer & Rock, 1994). Analogously, sounds that are presented closer together in time are more likely to form a single perceptual unit or stream (Handel, Weaver, & Lawson, 1983). A logical consequence of

the law of proximity is that sounds that are far apart in time, and visual stimuli that are far apart in space, will fail to form perceptual units (Bregman, 1990). For example, previous research has indicated that sounds presented more than 1.8 – 2 s apart are not perceived as part of the same stream of sounds (Mates, Müller, Radil, & Pöppel, 1994) and that the visual system fails to group objects together as the space between them increases (Palmer & Rock, 1994).

Recently, Baker et al. (2004) examined the impact of spatial perceptual grouping on visual ISL. Participants were presented with statistical patterns of simultaneously presented pairs of visual shapes; pairs were either spatially connected by a bar (a strong form of visual perceptual grouping) or not. They found that participants in the stronger perceptual grouping condition had better learning than those in the weaker perceptual grouping conditions. Similar results have been found by Pacton and Perruchet (2008). These studies demonstrate that spatial perceptual grouping conditions affect visual ISL.

To date, the relationship between perceptual grouping and learning in the auditory modality has not been systematically investigated. If strong perceptual grouping aids ISL, then auditory perceptual grouping ought to improve as sounds are presented at closer temporal proximity (i.e., at a faster rate). Conway and Christiansen (2009) reported that increasing rates of presentation from 4 stimuli/second (250-ms stimulus onset asynchrony, SOA) to 8 stimuli/second (125-ms SOA) did not impact learning in the auditory modality. However, this is a small range of presentation rates, with both rates being well within the limits of auditory perceptual grouping (i.e., SOA less than 2 s). In order to more directly assess the effects of temporal perceptual grouping, more varied grouping conditions need to be examined for both auditory and visual input.

3.1.3 Current experiments

The current paper examines the effect of perceptual grouping along the temporal dimension using greater changes in presentation rate than have been previously investigated. Specifically, the current experiment examines both visual and auditory ISL when the streams are presented either at fast rates of presentation (similar to rates used in previous studies) or under much slower rates of presentation. If auditory ISL is aided by temporal perceptual grouping, auditory learning should improve when sounds are presented closer together in time (i.e., at a faster rate) and should be disrupted when sounds are presented further apart in time (i.e., at a slower rate).

In contrast, we predict the opposite effect of presentation rate on visual ISL: Since visual processing has poorer temporal resolution, visual ISL should not be facilitated by a fast rate of presentation as auditory ISL would. Instead, visual ISL will improve with slower rates of presentation because this is less temporally demanding on the visual system. Previous work has demonstrated improvements to visual ISL with slower rates of presentation (Conway & Christiansen, 2009; Turk-Browne et al., 2005).

It is crucial to note that the changes in temporal rate employed in the current study do not obfuscate the individual stimuli themselves. At the fastest rate of presentation employed in the current study, previous work (Conway & Christiansen, 2005) as well as pilot testing revealed that there is robust perception of individual visual and auditory stimuli. Thus, by "changes in perceptual conditions" we are not referring to changing the ability of participants to perceive individual stimuli. However, as reviewed above, changes in rate of presentation have been shown to affect perception of auditory stimuli as occurring in a

single stream and to decrease ability of the visual system to resolve streams of stimuli. Thus, it is the perception of these streams of stimuli in which statistical regularities are presented, but not the individual stimuli that is being affected by differences in rate of presentation.

In the current paradigm, participants are familiarized with both visual and auditory statistical regularities. Conway and Christiansen (2006) observed that statistical information from two different streams could be learned simultaneously if these streams were from different modalities (visual and auditory) but not if they were instantiated in perceptually similar stimuli. In their design, strings of stimuli were generated by two different artificial grammars and interleaved with one another, as complete strings, in random order. In the current study, we investigated statistical learning of triplets of stimuli within a single stream (Figure 3.1A). Since triplet boundaries are key statistical information, alternating between full triplets would provide an explicit boundary cue. To avoid such a scenario while presenting both auditory and visual triplets, we adapted the interleaved design from Turk-Browne et al. (2005) to present an auditory and a visual familiarization stream (see Figure 3.1B for illustration of the interleaved design as applied to the current study). In addition, interleaving two familiarization streams avoids cross-modal effects in ISL that have been observed when visual and auditory streams are presented simultaneously (Robinson & Sloutsky, 2007).

Thus, if ISL is affected by modality-specific or perceptual processes, we predict that rate manipulations will have opposite effects on visual and auditory ISL: (a) We expect auditory ISL to be poorer at slower rates of presentation than learning at fast rates, and (b) we predict the opposite pattern of results in the

visual modality: We expect learning to be stronger at slow presentation rates and weaker at fast presentation rates.

In addition to manipulating the rate of presentation in the current study, we also manipulate selective attention to the streams. While the necessity of attention is unclear in ISL (Saffran et al., 1997), it has recently been established that selective attention to the information containing the statistical regularities boosts performance in both the visual and the auditory modalities (Toro et al., 2005; Turk-Browne et al., 2005). Consistent with this work, we predict that there will be significantly reduced learning for the unattended streams for both visual and auditory sensory modalities with both rates of presentation. Thus, we do not expect to see an effect of rate in the unattended streams given that we anticipate seeing no learning in conditions without attention.

Focusing on predictions for the attended streams, it has been proposed that one way in which attention aids in ISL is through boosting performance when perceptual grouping conditions are unfavorable. Recent work has suggested that poor perceptual grouping conditions can be overcome with selective attention to relevant stimuli (Baker et al., 2004; Pacton & Perruchet, 2008). However, the type and range of perceptual grouping in these studies has been limited, and investigations have not extended beyond the visual modality. It is unknown whether selective attention can overcome poor grouping conditions in the auditory modality and whether attention is always sufficient to overcome even extreme disruptions in perceptual grouping.

Given the large variations in temporal rate in the current studies, we predict that selective attention will not be sufficient to compensate for the poor perceptual conditions induced by these changes in presentation rate. Thus, we expect

to see that the modality-specific effect of temporal rate (i.e., poor at fast rates for visual and poor at slow rates for auditory) will persist even if participants selectively attend to these modalities. An interaction of rate and modality under conditions of selective attention would be evidence that selective attention is not always sufficient to compensate for poor perceptual conditions.

3.2 Experiment 1: Interleaved, fast presentation (375-ms SOA)

To examine the modality-specific effects of temporal perceptual grouping (rate of presentation), we interleaved two familiarization streams governed by statistical information in the visual and auditory modalities. The current experiment presented streams at a rate similar to that in previous ISL studies (SOA less than 500 ms). As with this previous work, we predict an auditory superiority effect in ISL at these relatively fast rates of presentation (Conway & Christiansen, 2005, 2009; Saffran, 2002).

Two familiarization streams (auditory and visual) were interleaved to create a single stream; this was done by sampling one to six elements at a time from a single stream consecutively (see Figure 3.1B). Interleaving streams resulted in a predictable set of transitional probabilities that was roughly equal across experimental groups (Table 3.1). Transitional probabilities are higher for successive elements within triplets than for those spanning triplets, providing a cue for learning (e.g., see Fiser & Aslin, 2001; Saffran et al., 1996; Turk-Browne et al., 2005).

As with Turk-Browne et al. (2005), selective attention was manipulated between streams. While some research has indicated that explicit attention to stim-

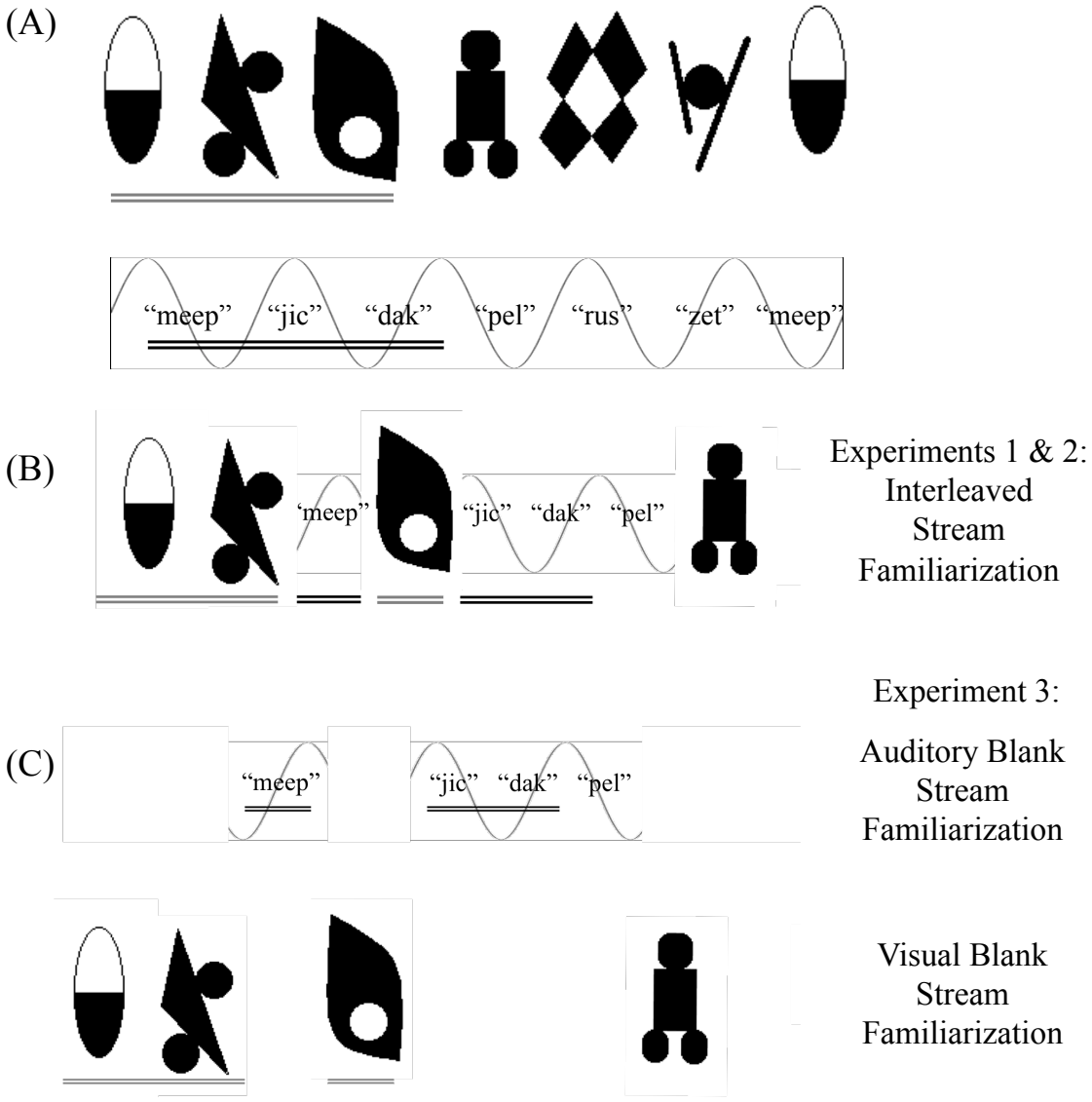


Figure 3.1: (A) A sample of separate visual and auditory familiarization streams prior to interleaving. Stimuli are presented sequentially with order of presentation from left to right. A triplet is underlined in each stream (visual: grey; auditory: black). (B) In Experiments 1 and 2, visual and auditory streams were interleaved such that stimuli from both modalities were presented sequentially with presentation pseudorandomly switching between streams with no more than six consecutive elements from a single modality. (C) In Experiment 3, interleaved streams were presented with the same timing of presentation for a stream from an attended modality but with unattended stimuli from the other modality removed.

uli is not required for ISL (Saffran et al., 1997), other research has demonstrated that selective attention aids in ISL in both the visual (Turk-Browne et al., 2005), and the auditory (Toro et al., 2005) modalities. Thus, we do not expect to see evidence of learning in unattended streams regardless of rate of presentation.

3.2.1 Methods

Participants

Thirty-two participants were recruited from psychology classes at Cornell University, earning extra credit or $10/hour. All participants reported normal or corrected-to-normal vision and no serious auditory deficits or neurological problems.

Materials

Auditory and visual stimuli were presented at a rate similar to that in previous statistical learning studies (e.g., Conway & Christiansen, 2005; Saffran et al., 1996, 1997): Visual and auditory stimuli are presented for 225 ms with an interstimulus interval (ISI) of 150 ms, resulting in an SOA of 375 ms. All stimuli were presented using E-prime stimuli presentation software (Version 1, Psychology Software Tools).

Visual stimuli. Fifteen novel abstract shapes were drawn using MS Paint for Windows 98 Second Edition (see Appendix A). The stimuli were designed to be perceptually distinct and not easily labelled verbally. During central presentation, shapes measured 4 cm by 6 cm on a 17- inch Samsung SyncMaster 955DF.

Participants were seated 65 cm from the screen.

Auditory stimuli. Fifteen monosyllabic nonwords, recorded by a female, native English speaker, were chosen to obey the phonological rules of English and be easily distinguishable from each other but as unique and unfamiliar as possible (see Appendix B). All nonwords were edited using Audacity for OSX (Version 1.2.2, Free Software Foundation, Boston, MA; Audacity Team, 2005).

Procedure

Participants were randomly assigned to one of three groups: two experimental groups, *visual attention* or *auditory attention* (24 participants), or *nonfamiliarized controls*. Participants in the two experimental groups had identical procedures except for the inclusion in the instructions that participants preferentially attend to a single modality. [2] Immediately following familiarization, experimental participants were tested for evidence of learning in both the visual and the auditory modalities. Participants in the *nonfamiliarized control* group were given the same testing procedure as those in the experimental condition but without receiving familiarization.

Familiarization. Stimuli were grouped offline into single-modality triplets resulting in five auditory and five visual triplets. In order to ameliorate any ef-

[2]Before familiarization, participants were instructed to attend to a single modality (auditory or visual) depending on their assigned group. They were instructed that stimuli in the other modality were meant to provide distraction. Participants were told to respond to the repeated elements in their assigned modality only. If participants were in the auditory attention group, they were specifically instructed to still look at the monitor but to just direct their attention to the auditory stimuli. Due to a data collection error, repeat responses were not collected. However, the replication of these results in Experiment 3 without unattended stimuli indicates (a) that participants are in fact attending to the assigned sensory modality and (b) that attention to a particular modality was analogous to attention during exposure without unattended stimuli (i.e., there was no interference).

fects of triplet grouping, multiple groupings were used across participants with each triplet grouping employed in all conditions. Thirty presentations of each triplet were randomly ordered such that no triplet or pairs of triplets were immediately repeated (e.g., ABCABC or ABCDEFABCDEF). A cover task was employed: Participants were asked to detect repeated elements in the familiarization stream using a button box, and no feedback was given. The first and third elements of each triplet were repeated two times during familiarization (e.g., ABCCDEFGGHI; Turk-Browne et al., 2005).

Auditory and visual familiarization streams were pseudorandomly interleaved by sampling each stream in order and without replacement with no more than 6 elements from one stream sampled consecutively (see Figure 3.1B). Critically, the process of interleaving did not highlight the triplet structure of the familiarization streams, with streams often switching between modalities within triplets. This resulted in a familiarization stream of 940 elements: 470 from each modality. Participants were given a self-timed break halfway through familiarization. The sequence of interleaving was counterbalanced such that the interleaved order of the visual elements for one group of participants was that of the auditory elements for another group of participants; attention was counterbalanced across modality and interleaved order.

Testing. Test trials were constructed for each modality separately comparing triplets from familiarization to foils (Figure 3.1A). Then test trials from both visual and auditory test trials were presented in random order in a multimodal testing block. Within each modality, the testing phase consisted of a forced-choice task pairing the five triplets constructed for each participant with five foils and counterbalanced for order of presentation, resulting in 50 test trials

Table 3.1: Transitional probabilities of elements in the stream for each modality in isolation and interleaved

	Isolation	*Interleaved*	*Isolation*	*Interleaved*
p(any particular shape), e.g., *p*(B)	1/5 x 1/3	1/15 x 1/2	.064	.032
p(any pair within a triplet), e.g., *p*(A, B)	1/15 x 1/1	1/30 x 1/2 x 1/1	.064	.016
p(any pair spanning triplets), e.g., *p*(C, G)	1/15 x 1/4	1/30 x 1/2 x 1/4	.016	.004
p(any given triplet), e.g., *p*(A, B, C)	1/15 x 1/1 x 1/1	1/30 x 1/2 x 1/1 x 1/2 x 1/1	.064	.008
p(any given nontriplet), e.g., *p*(B, C, G)	1/15 x 1/1 x 1/4	1/30 x 1/2 x 1/1 x 1/2 x 1/4	.016	.004
p(any foil sequence), e.g., *p*(A, B, I)	0	0	0	0

per modality (5 triplets x 5 foils x 2 order). The same foils were paired with all triplets during test; thus there were the same number of foils and triplets used at test to equate exposure. Foils were constructed from the same shapes and nonwords, designed to violate the triplet structure but not absolute element position (e.g., triplet: ABC, DEF, GHI; foil: ABF, DEI, GHC). All of these stimuli were presented in the same manner and with the same timing as the familiarization stream. Foils and triplets were separated by 1,000 ms of silence. Following the methodology of Conway and Christiansen (2005) and Saffran (2002), participants were instructed to report which triplet seemed "more familiar or right based on [their] previous task, if applicable". They were instructed not to respond based on the familiarity of the individual elements. After presentation of a pair of test items, participants were prompted to press Key 1 (of a 4-key response pad) if they felt that the first item was more "familiar" or "right" and to press Key 4 for the second item. The response screen was self-timed and participants received no feedback on their responses. Participants were instructed that there was no order to the modality of successive test trials. The dependent measure was accuracy in discriminating triplets from foils across 50 test trials.

3.2.2 Results

Results are collapsed across both interleaved pattern and triplet groupings with analysis occurring only along dimensions of experimental groups (auditory vs. visual attention) and experimental versus nonfamiliarized controls.

Nonfamiliarized Controls

Performance of participants in the control group was evaluated against chance performance (25 out of 50, or 50%). Control participants performed at 49% accuracy for both modalities, and neither was significantly different from chance performance: visual, $t(7) = 0.36$, $p = .73$; auditory, $t(7) = 0.80$, $p = .45$.

Experimental groups

Participants who attended to auditory stimuli correctly responded to 63% of auditory test trials and 54% of visual test trials. Those who attended to visual stimuli correctly responded to 57% of visual test trials and 47% of auditory test trials (see Figure 3.2). Comparing experimental performance to corresponding control performance, only the attended auditory condition differed significantly from nonfamiliarized controls, $t(18) = 5.95$, $p < .001$; auditory unattended, $t(18) = 0.420$, $p > .5$; visual attended: $t(18) = 1.73$, $p = .10$; visual unattended: $t(18) = 1.336$, $p = .20$.

Effects of attention. To specifically investigate the effects of selective attention in the interleaved-multimodal design, planned t tests were performed to compare performance for a single modality in attended and unattended conditions, across experimental groups. This comparison of attended and unattended streams yielded a significant difference in the auditory modality only: auditory attended versus unattended, $t(22) = 4.16$, $p < .01$; visual attended versus unattended, $t(22) = 0.90$, $p = .38$.

Modality effects. Experimental data were submitted to a two-way analysis of variance (ANOVA; visual vs. auditory attention, within-subject factor: visual

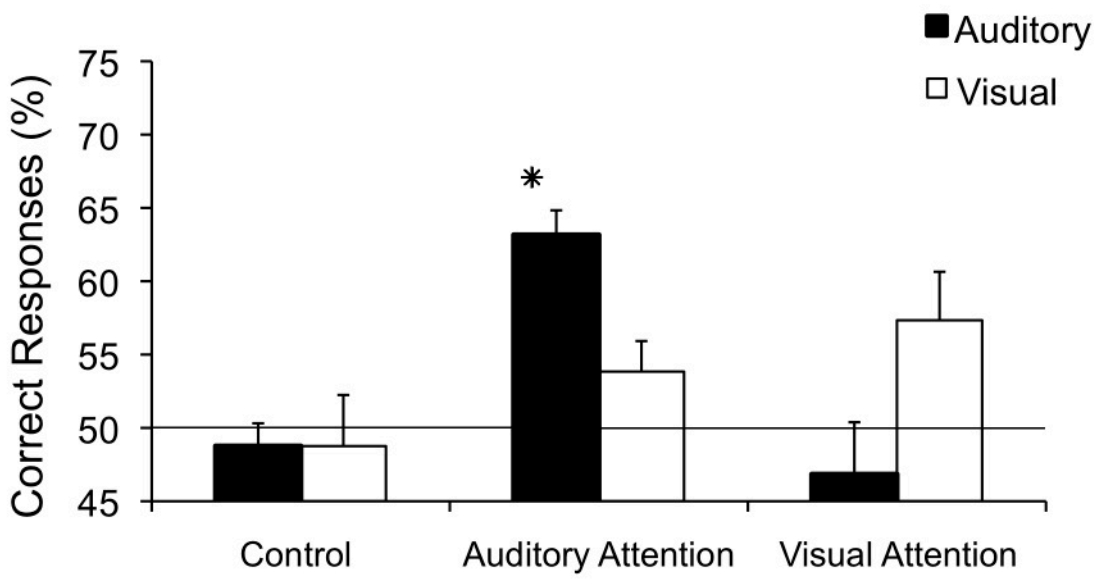


Figure 3.2: Mean test performance (percentage correct out of 50) from Experiment 1. Visual and auditory ISL (implicit statistical learning) performance is presented for control, auditory and visual attention conditions at fast presentation rate (375-ms stimulus onset asynchrony, SOA).

vs. auditory presentation). While there is no main effect of modality, $F(1, 22)$ 1= 0.056, p $>$ .5, there is a significant modality by attention interaction, $F(1, 22) = 16.21$, $p = .001$. That is, modality effects were obtained specifically when participants were devoting attention to a given input stream. While direct tests of attended performance across modalities do not reveal a significant difference, $t(22) = 1.573$, $p > .1$, the interaction of modality and attention indicates that modality of presentation is not uniformly affecting learning across attentional conditions. Together with the results presented earlier (specifically, a significant effect of attention in the auditory modality only and significant learning is restricted to the attended auditory stream), this interaction indicates that auditory ISL is superior to visual ISL at this rate of presentation when selective attention is deployed. Increased ISL in the auditory modality is consistent with previous findings using similarly timed rates of presentation (e.g., Conway &

Christiansen, 2005).

3.2.3 Discussion

Here we used a multimodal interleaved design to investigate auditory and visual ISL. This experimental design is a novel combination and extension of that used by Conway and Christiansen (2006) and Turk-Browne et al. (2005). Our results corroborate previous cross-modal ISL findings. First, using similar rates of presentation in the current study, auditory ISL appears to have superior performance to visual ISL (Conway & Christiansen, 2005; Robinson & Sloutsky, 2007; Saffran, 2002). Second, concerning the effect of attention, our results are again consistent with previous studies showing that attention can improve learning (Toro et al., 2005; Turk-Browne et al., 2005). However, we found a significant interaction that, together with follow up t-tests, indicated that selective attention improved auditory learning more than visual learning. Thus, at this relatively fast presentation rate, only auditory learning occurred, even when selective attention was available. Under the same presentation conditions, we do not find evidence of visual learning even with the aid of selective attention. This is likely because, while individual stimuli are easily perceived at the current rate of presentation, visual processing has relatively poor temporal resolution in the current task. See the introduction for a more in-depth discussion.

3.3 Experiment 2: Interleaved, slow presentation (750-ms SOA)

The results from Experiment 1 are consistent with those from previous studies demonstrating superior auditory learning at fast presentation rates (when the input is attended). In the current experiment, we move beyond the temporal distances previously explored in the ISL literature by increasing the distance between successive elements from 375-ms SOA to 750-ms SOA, effectively increasing the amount of time between successive elements in the presentation stream. In fact, given the interleaved design and the increased rate of presentation, the average amount of time between successive visual-to-visual or auditory-to-auditory elements is 2.25 s.[3] Thus, this rate of presentation provides input conditions that are beyond the perceptual grouping tolerance of the auditory system (Mates et al., 1994). See Figure 3.3 for an illustration of the relative length of pauses for a single element (average is three elements) in Experiment 1 (top panel) and Experiment 2 (centre panel) relative to the length of pause necessary to produce significant temporal grouping disruption (bottom panel).

As addressed above, this slower rate should have opposite effects on visual and auditory ISL. Given that weak spatial perceptual grouping can reduce visual ISL (Baker et al., 2004), we predict a similarly negative effect for weak temporal perceptual grouping on auditory ISL. Thus, we predict that a slow rate of presentation will have a negative effect on auditory ISL. However, given that the visual system has relatively poor temporal processing in the current task

[3] In the current experimental methods, there were between one and six stimuli from a single familiarization stream presented consecutively. The mean number of consecutive stimuli was three, which, at the rate of presentation employed in Experiment 2, has a duration of 2.25 s. Thus, the average length of pause in an attended familiarization stream, caused by presentation of the unattended familiarization stream, was 2.25 s.

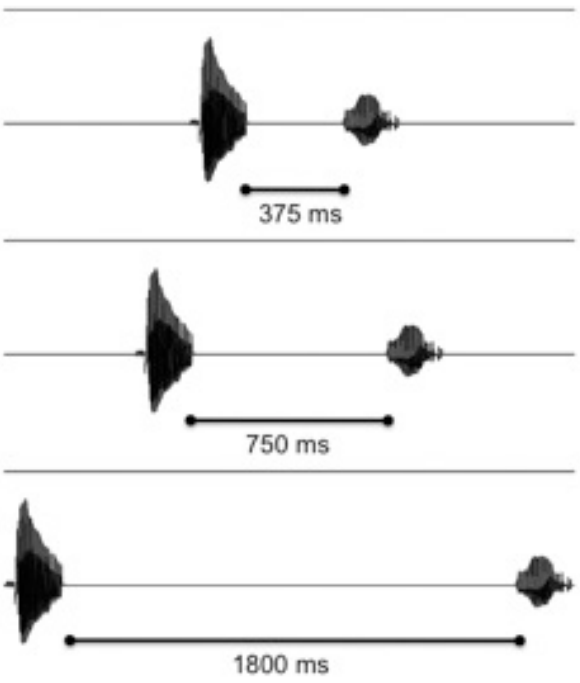


Figure 3.3: Illustration of the temporal separation created by the interleaving of a single unattended element at the fast (375-ms stimulus onset asynchrony, SOA) and slow (750-ms SOA) presentation speeds in relation to the limits of auditory temporal perceptual grouping (1.8 – 2 s). On average, three unattended elements were presented consecutively.

(see Footnote 1), a decreased rate of presentation should have a positive effect on visual ISL because it places less of a demand on the visual system than does the fast presentation rate used in Experiment 1.

3.3.1 Methods

All methods, materials, and procedures were identical to those in Experiment 1 with the exception of presentation rate: Both visual and auditory stimuli were present for 450 ms with a 300-ms ISI (750-ms SOA). To accommodate a slower rate while maintaining natural production, a largely overlapping set of mono-

syllabic nonwords were recorded by a female, native-English speaker (see Appendix B) and were edited to 750-ms SOA.

Another 32 participants were randomly assigned to one of three groups: two experimental groups (*visual attention* or *auditory attention*), or *nonfamiliarized controls*.

3.3.2 Results

Nonfamiliarized controls responded correctly to 43% of the visual test trials and 46% of the auditory test trials; neither result was significantly different from control performance at 50%: visual, $t(7) = 1.27$, $p = .25$; auditory, $t(7) = 1.42$, $p = .20$. Participants who attended to the visual modality correctly responded to 65% of visual test trials and 48% of auditory test trials. Those who attended to the auditory modality correctly responded to 55% of auditory test trials and 52% of visual test trials (Figure 3.4). In contrast to the results from Experiment 1 (see Figure 3.2), only performance on the attended visual stream was significantly different from control performance: attended visual, $t(18) = 3.67$, $p = .002$; unattended visual learning, $t(18) = 1.73$, $p = .10$; attended auditory, $t(18) = 1.81$, $p = .087$; unattended auditory, $t(18) = 0.85$, $p = .85$.

Effects of attention

Planned comparison of attended versus unattended performance within modality yielded a significant difference in the visual modality only: attended versus unattended visual, $t(22) = 3.44$, $p = .002$; auditory, $t(22) = 1.65$, $p = .105$.

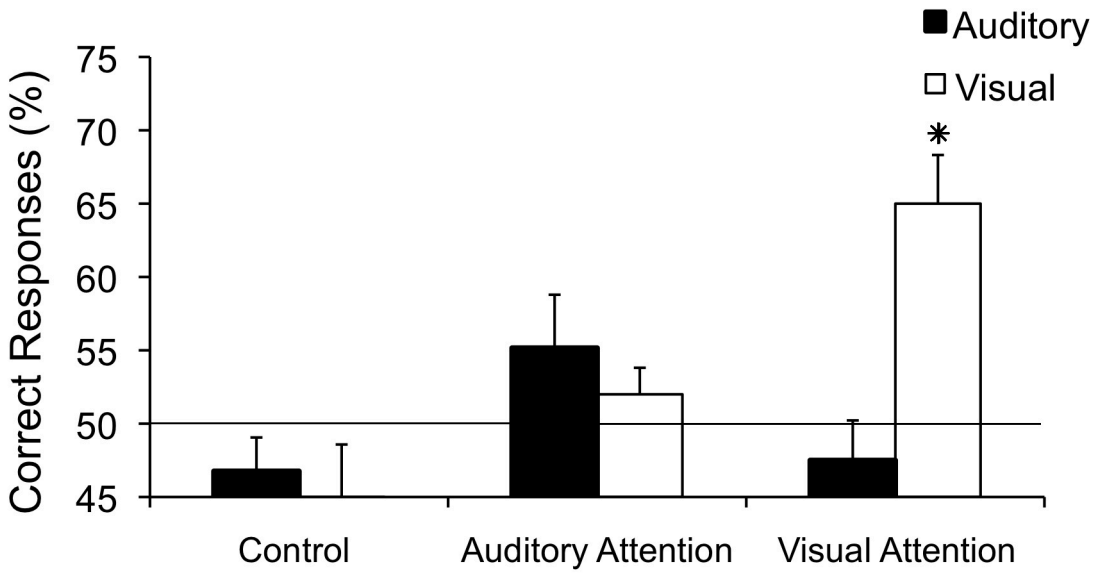


Figure 3.4: Mean test performance (percentage correct out of 50) from Experiment 2. Visual and auditory ISL (implicit statistical learning) performance is presented for control, auditory and visual attention conditions at fast presentation rate (750-ms stimulus onset asynchrony, SOA).

Modality effects at slower rates of presentation

As with the results in Experiment 1, data from the experimental groups were submitted to a repeated measures ANOVA. We find a main effect of modality, $F(1, 22) = 4.32$, $p = .050$. This main effect is supported by greater performance in the visual modality overall. However, as reported above, there is no evidence of unattended learning in the visual modality suggesting that the main effect is driven by attended visual performance.

As in Experiment 1, we found a significant modality by attention interaction, $F(1, 22) = 8.98$, $p = .0066$. Direct t-tests of attended results across modalities revealed only a marginally significant difference, $t(22) = 2.0$, $p = .058$. However, taken with the t-tests comparing experimental performance to controls, the in-

teraction of modality and attention indicates that modality of presentation is not uniformly affecting learning across attentional conditions.

Unlike Experiment 1, direct t-tests indicate an opposite pattern of results of attention and learning: We find a significant effect of attention in the visual modality only and that significant learning is restricted to the attended visual stream. These results suggest that at the slower rate of presentation, visual ISL performed more robustly than auditory ISL. This is in contrast to the auditory superiority in ISL at fast presentation rates using the same materials and relative temporal dynamics as those in Experiment 1.

Comparing across rates of presentation

A direct comparison of results from Experiments 1 and 2 was conducted by submitting all experimental data to a three-way repeated measures ANOVA using between-subject factors of attended modality (auditory vs. visual) and rate (fast vs. slow) and within-subject factor of modality (auditory vs. visual). This analysis revealed a marginally significant main effect of modality, $F(1, 44) = 3.35$, $p = .074$, driven by slightly better performance across attended and unattended streams in the visual modality. In addition, we find a significant interaction of modality by attended modality, $F(1, 44) = 22.84$, $p < .002$. This interaction confirms that attention to a particular modality affects performance in the corresponding modality for both the visual and the auditory modalities. We also find a significant interaction between attended modality and rate of presentation, $F(1, 44) = 5.58$, $p = .023$. This result shows that the effect of rate is dependent on which modality is being attended. We do not obtain a rate by modality interaction, $F(1, 44) = 2.414$, $p = .127$, nor a modality by rate by attended modality

interaction, $F(1, 44) = 0.010$, $p = .922$, which indicates that the rate manipulation does not affect one modality preferentially over the other (except when the modality is attended).

Taken together, the interactions of attended modality with rate and with modality of presentation provide support for our hypothesis that modality of presentation mediates differential performance across rates of presentation but this only occurs in the attended streams. This view is supported by the pattern of significant learning (i.e., greater performance than that of controls) in Experiments 1 and 2.

These results are broadly consistent with previous work that attention modulates ISL. However, they reveal a more complex relationship between attention and learning: Attending to a specific modality does in fact have an effect on ISL only in that modality–moreover, this modality-specific effect of attention interacts with rate of presentation. This pattern of results suggests that attention is not sufficient to overcome the modality-specific effects of rate.

Given the broad and complex effects of attention in the current task and our specific predictions that rate will interact with learning in the attended streams (see the introduction), we conducted separate analyses for attended and unattended performance to examine the effects of modality and rate within attentional condition. This analysis allows us to verify that the interactions discussed above are in fact driven by attended performance and not patterns of unattended learning. For each group (attended and unattended), we submitted the data to a repeated measures ANOVA (rate of presentation and within-subject factor: modality). For the unattended conditions, we found a main effect of modality, $F(1, 44) = 4.80$, $p = .034$, but no effects of rate or interaction of rate

and modality. Examination of the mean unattended performance across conditions reveals that visual unattended learning is superior to auditory unattended learning. These results confirm that there is no effect of rate in unattended performance.

The same analysis of attended performance revealed a markedly different pattern of results. We find no main effect of modality or rate ($Fs < 1$) but a significant modality by rate interaction: $F(1, 44) = 6.47$, $p = .015$. These results confirm our predictions that modality and rate interact with learning but only when attention is deployed. These results also confirm that selective attention is not sufficient to overcome modality-specific differences in learning.

3.3.3 Discussion

The perceptual literature predicts that decreasing rate of presentation will have opposite effects on auditory and visual processing: A decrease in rate will disrupt auditory perceptual grouping while simultaneously easing the temporal processing in the visual modality. We find this differential pattern of performance in ISL, indicating that perceptual conditions significantly impact implicit statistical learning ability. In the current experiment, rate of presentation is slowed to half the rate of Experiment 1. While auditory superiority is observed at the "fast-rate", we fail to observe any significant learning in the auditory modality at the current "slow-rate" of presentation. Instead, we observe a visual superiority effect. This switch of ISL across perceptual modalities is particularly striking given that no other methodological changes were made.

Consistent with previous research (Toro et al., 2005; Turk-Browne et al.,

2005), we find that attention aids ISL. Additionally, we find that timing and attention interact. This finding runs counter to recent research that has suggested that attention is sufficient to overcome poor perceptual grouping conditions in the visual modality (Baker et al., 2004; Pacton & Perruchet, 2008). Instead, we find evidence that attention is not sufficient to overcome adverse perceptual grouping conditions. The next experiment further controls for the presence of unattended stimuli and any effects of cross-modal presentation in Experiment 2, in addition to providing a replication of the rate by modality interaction.

3.4 Experiment 3: Interleaved, blank stream

Finally, Experiment 3 addresses the possibility that the cross-modal interleaved design employed in Experiments 1 and 2 introduced cross-modal interference or decrements of selective attention due to the presentation of unattended stimuli from another modality (Spence & Driver, 1997). To this end, Experiments 3A (375-ms SOA) and 3B (750-ms SOA) use the same timings of the attended streams as those in Experiments 1 and 2, respectively, but, instead of presenting unattended stimuli from the second input stream, pauses of equivalent length were presented. For example, if three unattended elements at 375-ms SOA were presented in Experiment 1, a pause of 1,125 ms (375 ms x 3) was presented in place of these unattended elements, as illustrated in Figure 3.1C. This manipulation preserved the temporal structure of the familiarization stream while removing any potential cross-modal interference and cost of dual- modality presentation. In addition, this experiment is meant to ameliorate any effects of the attentional manipulations employed in Experiments 1 and 2 and increases transitional probabilities of the familiarization stream. Without the unattended

elements, transitional probabilities are equivalent to presenting the stream in isolation (see Table 3.1).

3.4.1 Methods

The attended streams from Experiments 1 and 2 were used in the current experiment with the unattended elements removed and replaced with equal-length pauses in stimulus presentation. Thus the timing, materials, and methods were preserved from the previous two experiments with the exception of the removal of the unattended, cross-modal stream.

Participants

Another 32 participants (16 each for Experiments 3A and 3B) were recruited from introductory psychology classes at Cornell University to participate in exchange for course credit or $10/hour.

Procedure

For each experiment, participants were randomly assigned to one of two conditions: nonfamiliarized controls or experimental groups. After being familiarized with the stimuli from one modality (with the other one "blanked out"), participants were tested in that same modality with test trials for that modality only presented in random order. Then, participants went on to the familiarization, as described above, and testing trials for the other modality. Test trials were constructed using the same procedure as that in Experiments 1 and 2. Modal-

ity order was counterbalanced across participants. The same number of triplet groupings and interleaved sequence were employed. Because a single unimodal familiarization stream was presented during familiarization, there was no manipulation of selective attention.

3.4.2 Results and Discussion

Experiment 3A

Participants in the nonfamiliarized control group responded correctly to 52% of visual and 51% of auditory test trials and did not perform significantly better than chance in either modality: visual, $t(7) = 0.97$, $p > .36$; auditory, $t(7) = 0.10$, $p > .92$. Participants in the experimental group responded correctly to 51% of the visual test trials and 64% of the auditory test trials. Only performance in the auditory modality was significantly better than that of nonfamiliarized controls: visual, $t(14) = 0.84$, $p = .41$; auditory, $t(14) = 2.33$, $p = .035$ (see Figure 3.5). We find a significant difference in mean performance across modalities, $t(15) = 4.79$, $p < .001$.

To verify that there were no effects of our within-subject design, we investigated possible order effects. Comparison of performance dependent on order of presentation revealed no effect of presentation order in either modality: visual, $t(6) = 0.131$, p $>.90$; auditory, $t(6) = 1.59$, $p = .16$.

Inspection of the left panel of Figure 3.5 readily reveals the replication of attended performance in Experiment 1 (see Figure 3.2). Statistical comparison between current experimental results and performance in attended conditions

from Experiment 1 revealed no significant difference in either modality: visual, $t(14) = 0.83$, $p = .42$; auditory, $t(14) = 0.18$, $p = .86$. Thus, current results replicate attended performance in Experiment 1, demonstrating that the presence of unattended cross-modal stimuli has no effects on ISL performance in either the visual or the auditory modalities.

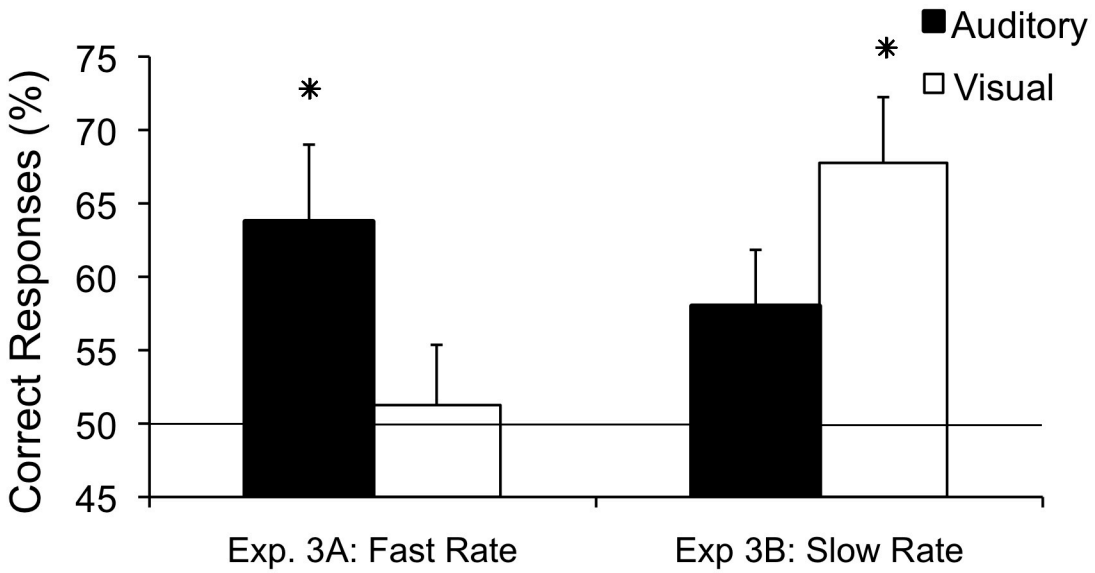


Figure 3.5: Mean test performance for Experiment 3. Auditory and visual streams are presented with identical timing to that in Experiments 1 and 2 but without the unattended stimuli. Both modalities are attended and presented in counterbalanced order within participants. Left: Experiment 3A using the fast rate of presentation from Experiment 1. Right: Experiment 3B using the slow rate of presentation from Experiment 2.

Together, the results from Experiments 1 and 3A demonstrate that auditory presentation yields greater learning performance than visual presentation at fast rates. This finding is consistent with previous studies, all of which have employed similarly fast rates of presentation (Conway & Christiansen, 2005, 2009; Robinson & Sloutsky, 2007; Saffran, 2002).

These results also suggest that auditory ISL is robust to short pauses in pre-

sentation. The interleaved experimental design entails that a single stream is temporally interrupted either by unattended visual stimuli in Experiment 1or pauses of equivalent length in Experiment 3A. At the fast presentation rate, these pauses average 1.13 s in length. Even with these short pauses, there is significant learning in the attended auditory stream . Learning remains intact likely because the length of pauses is on average within the perceptual grouping tolerance of the auditory modality (Mates et al., 1994; see introduction).

Experiment 3B

Nonfamiliarized controls responded correctly to 51% of visual and auditory test trials and did not perform significantly better than chance in either modality: visual, $t(7) = 0.39$, $p = .71$; auditory, $t(7) = 0.56$, $p = .59$. Participants in the experimental group responded correctly to 68% of the visual test trials and 58% of the auditory test trials. In contrast to results from Experiment 3A, only performance in the visual modality differed significantly from that of nonfamiliarized controls: visual, $t(14) = 2.30$, $p = .037$; auditory, $t(14) = 1.66$, $p = .12$ (see Figure 3.5). We do not, however, find a significant difference in learning across modalities, $t < 1$.

As in Experiment 3A, we tested for possible order effects by comparing performance in each modality dependent on presentation order. We found no evidence of order effects for either modality: visual, $t(6) = 0.82$, $p = .45$; auditory, $t(6) = 0.62$, $p = .56$.

Since the purpose of the current experiment was to replicate the results from Experiment 2 (see Figure 3.4), performance in the attended streams from Experi-

ment 2 was compared to the current experimental group. This analysis revealed no significant difference in either modality: visual, $t(14) = 0.738$, $p = .47$; auditory, $t(14) = 0.54$, $p = .6$ (see Figure 3.5). Thus, along with Experiment 3A, we find no effect of the presence of unattended cross-modal stimuli on ISL across presentation rates.

Analyses comparing attended performance in Experiments 1 and 2 suggest that timing and modality interaction in ISL. To test for this pattern of results in the within-subject design of Experiments 3A and 3B, data from experimental groups were submitted to a repeated measures ANOVA (auditory and visual ISL; timing). This analysis revealed no main effect of modality, $F(1, 14) = 0.097$, $p = .76$. Crucially, we find a significant modality by timing interaction, $F(1, 14) = 6.36$, $p = .024$. These results confirm the earlier result that presentation rate has a differential effect on ISL across visual and auditory modalities in two different experimental paradigms. [4]

3.5 General Discussion

In three experiments, participants were presented with auditory and visual statistical regularities under different timing conditions. While statistical information remained the same, performance was not equivalent across perceptual modalities at different rates of presentation. In attended modalities, perceptual modality and rate of presentation interact to affect learning: At fast rates of presentation, similar to previous studies, auditory ISL performed better than

[4]While the significant interactions of modality and timing or rate of presentation in Experiments 3A and 3B replicate the effects found in Experiments 1 and 2, in both of these experiments the interaction of rate and modality is found across participants and without random sampling across these experimental groups (i.e., there is only random sampling within Experiment 1).

visual; however, at a slower presentation rate, auditory ISL was reduced, and visual ISL became superior. This effect is replicated in Experiment 3, in which the timing conditions are kept constant but streams are presented without unattended stimuli. Thus, the current experiments have uncovered a new phenomenon in ISL where timing changes result in differential effects on ISL across visual and auditory modalities.

These learning results are consistent with well-known perceptual differences across visual and auditory modalities. As reviewed in the introduction, visual and auditory modalities appear to process time and space differently. Specifically, perceptual performance across modalities can be characterized by the visual:spatial::auditory:temporal analogy, where the visual modality preferentially processes spatially arrayed information while the auditory modality processes temporal information more robustly than the visual modality. While it is clear that under some circumstances this characterization does not apply (e.g., Potter, 1976), in the current type of perceptual task, these modality-differences are well established. While this perceptual effect does not limit the ability to recognize single objects, spatial or temporal presentation of multiple stimuli does affect processing across visual and auditory sensory modalities.

Our learning results parallel this well-known effect in the perceptual literature: Visual learning is increased with a slower rate while auditory processing is decreased. This novel phenomenon along with previous studies (e.g., Conway & Christiansen, 2005, 2006; Robinson & Sloutsky, 2007; Saffran, 2002) have established that ISL is not equivalent in the visual and auditory sensory modalities, even though the same statistical information is presented.

We also investigated the effect of attention to a sensory modality across both

timing conditions. While it is generally accepted that attention is a significant modulatory factor aiding in both visual and auditory ISL (visual: Turk-Browne et al., 2005 auditory: Toro et al., 2005), it is unclear whether attention is necessary for learning to take place (e.g., Saffran et al., 1997). We do not find any evidence for unattended learning (in Experiments 1 and 2). Therefore, we corroborate previous research showing that attention significantly boosts ISL in both visual and auditory sensory modalities and may be necessary for learning.

Turning to results for attended modalities, previous research on perceptual grouping effects in ISL has emphasized the relationship between perceptual conditions and attention. Baker et al. (2004) point to the "automatic spreading of attention induced by perceptual-grouping" (p. 465) as the mechanism by which perceptual grouping aids ISL. Under unfavourable grouping conditions, it has been argued that ISL can occur if the relevant stimuli are attended (Baker et al., 2004; Pacton & Perruchet, 2008). This approach emphasizes perceptual grouping as a factor that modulates attention, which in turn facilitates ISL, as opposed to treating perceptual grouping as a separate modality-specific factor affecting learning from statistical regularities.

If it were the case that attention is sufficient to compensate for poor perceptual conditions, we should observe equivalent learning in all attended streams regardless of modality and rate of presentation. The current results do not support this view: We observed a modality-specific decrement in ISL under disruptive perceptual grouping conditions even when there are no competing demands for attention. Thus, we find evidence that attention cannot always compensate for poor perceptual conditions.

Thus, our results suggest that favourable perceptual conditions and selective

attention to a particular modality may both be required in order for participants to learn from statistical regularities in the environment. In the current learning paradigm, attention appears to be a prerequisite for learning: There is no learning in any unattended stream regardless of perceptual modality or timing condition. However, selective attention to a particular modality is not sufficient for learning. In fact, even with attention, perceptual grouping conditions and modality of presentation interact to determine whether or not participants learn. Thus, we find that attention and perceptual conditions amenable to a particular modality are both necessary for ISL.

It is important to consider that each sense has unique sets of sensori-motor constraints that might affect learning and relevant for the current chapter, could affect learning in differently dependent on timing of sensory input. For example, eyes require foveation of relevant stimuli for the conveyance of detailed sensory input while ears do not have an apparent equivalent to a fovea. SImilarly, visual input is interrupted by the sensorimotor requirement of blinking and is interrupted during saccades (Burr, Morrone, & Ross, 1994), while, again, there is not obvious acoustic equivalent. During fast presentation of sensory input, it is possible that these sensori-motor constraints could reduce visual processing selectively in a way that is not dependent upon perceptual processing. There is some evidence that the decreased visual statistical learning at faster rates of presentation is not due to these sensori-motor constraints on the visual system. A control study was conducted to examine whether there were masking effects in the visual modality at faster rates of presentation. A masking effect would be evidenced by a selective decrease in discriminability of individual shapes at faster rates suggesting that the visual system is receiving reduced or compromised sensory input possibly as a result of these sensori-motor constraints. We find no

evidence of decreased discriminability in general at faster rates of presentation and no selective decrease in discriminability in the visual modality; the details of this experiment are presented in Appendix C. However, Turk-Browne et al. (2005) found that cover task performance was reduced in the visual modality at faster rates of presentation. Future research is needed to determine if a similar effect is found in the auditory modality. If there is an unequal reduction in cover task performance across perceptual modalities at fast rates of presentation, this could indicate that sensori-motor constraints rather than differences in perceptual processing, could be the source of the modality-specific effect of timing on statistical learning.

Overall, these results are consistent with ISL being mediated by mechanisms that are sensitive to the *perceptual nature* of the input in addition to its statistical structure (e.g., transitional probabilities and frequencies). Recent neuroimaging evidence has supported just such a scenario: Turk-Browne, Scholl, Chun, and Johnson (2009) report increased visual cortex activity during the observation of learnable visual sequences demonstrating that sensory cortices are likely involved in learning the underlying structure from visual statistics. Thus, both behavioural and neuroimaging results have suggested that ISL is sensitive to perceptual processes.

Given these modality-specific learning effects, any mechanisms for ISL must be able to account for differences in learning across sensory modalities. Finally, we consider three types of mechanisms for their ability to accommodate modality-specific differences as well as the domain-generality of ISL. In Figure 3.6, we present a simplified characterization of these possible architectures. Throughout the paper, we have been referring to ISL as a behavioural

phenomenon in which exposure to statistical regularities affects subsequent behaviour. By contrast, these candidate characterizations present types of possible mechanisms that could underlie this behaviour: Specifically, we consider different ways in which perceptual systems, responsible for the recognition and understanding of sensory information, and learning mechanisms, the acquisition of knowledge as a result of experience, could contribute to ISL ability.

As asserted in the introduction, it is difficult for the *standard view* of the mechanisms underlying ISL to account for these modality-specific patterns of learning. According to this view, ISL involves a single undifferentiated mechanism for which the nature of the input beyond its statistical characteristics is irrelevant (e.g., Perruchet & Pacton, 2006; Reber, 1989; Shanks et al., 1997). In other words, visual and auditory perception are separate unimodal processes that feed into a single learning mechanism. Consistent with this type of architecture, many prominent models make an "equivalence assumption" that the same statistical information presented across modalities should result in equivalent learning. Thus, while knowledge gained from ISL may be specific to the input stimuli (i.e., learning has limited or no transfer to other stimuli), the learning mechanism is not affected by the perceptual nature of the input (e.g., auditory stimuli are learned equivalently to visual stimuli). This standard architecture is consistent with domain-general nature of ISL given that there is a single learning mechanism that receives input from both perceptual systems. However, in our view, it is unclear how a learning mechanism that focuses on statistical information to the exclusion of the perceptual nature of the input would be able to accommodate the differential learning effects we report. One possibility is that the unimodal perceptual systems are tasked with detecting perceptual "units", while the singular learning mechanism can determine and track the statistical

regularities of these units, and differences in timing or other perceptual conditions make resolution of "units" in a stream of stimuli more difficult. However, the most obvious definition of a perceptual unit is a single object. In this case, in the current study and previous work, the perception of individual perceptual objects is the same across modalities.

We present two alternative types of architecture that, we believe, more easily accommodate the modality-specific patterns of learning observed in the current ISL paradigm. The first is a modality- specific architecture consisting of separate but computationally similar learning mechanisms for both visual and auditory perception, thus allowing perception and learning to remain distinct processes. The multiple, modality-specific learning mechanisms, which characterize this type of architecture, make it possible for differences in learning across sensory modalities to emerge in the learning mechanisms themselves (e.g., by detecting types of patterns similar to previous experience, spatial patterns in the visual system), as well as accounting for the modality- and domain-general nature of ISL. Two disadvantages to this architecture are that it is unparsimonious as it includes an additional, if computationally similar, learning mechanism and does not explain cross-modal learning effects (e.g., Robinson & Sloutsky, 2007).

In an *embodied* architecture, perception and learning are not distinct processes but the learning mechanism is a part of, or embodied in, perceptual processing. Recent work has suggested that perception is a prediction-based process (e.g., Summerfield, Trittschuh, Monti, Mesulam, & Egner, 2008) where the likelihood of a given stimulus affects perceptual processing. In this view of perception, sensitivity to statistical information is already a property of visual and auditory perception. Thus, changes in perception as a result of statistical in-

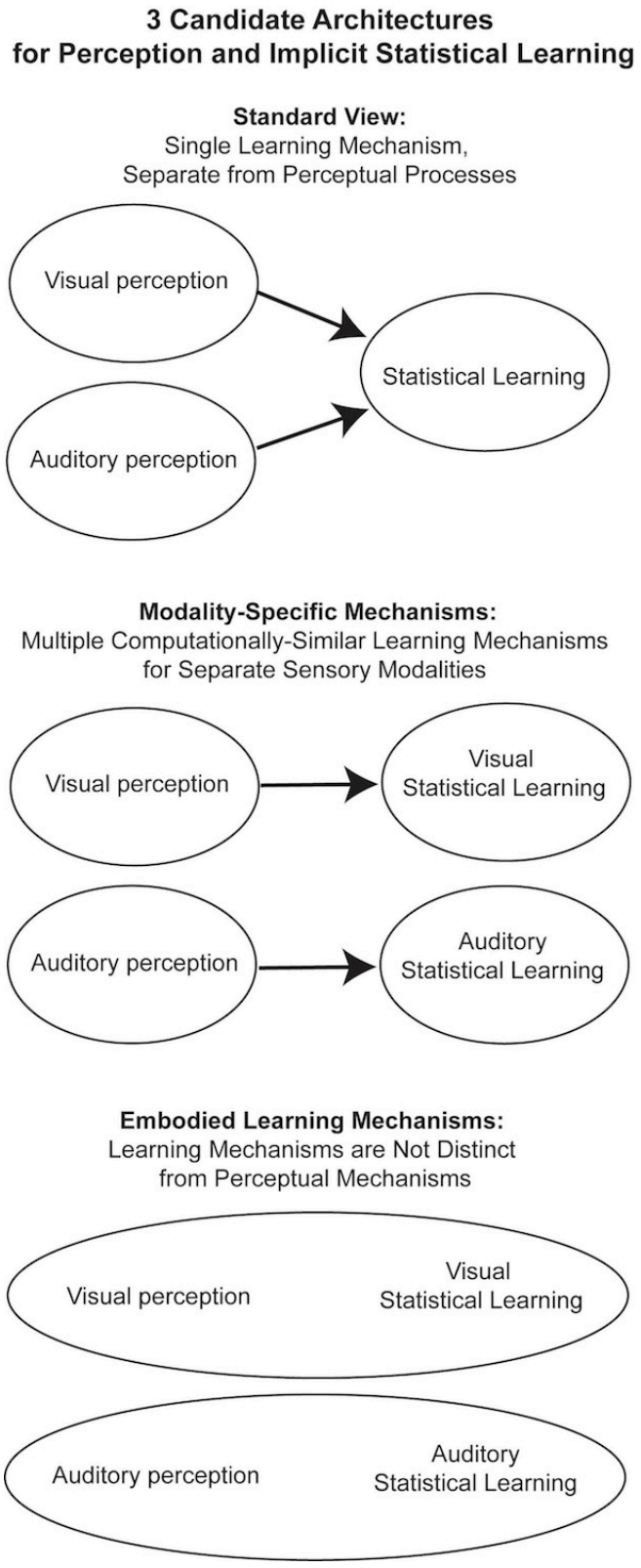


Figure 3.6: Simplified characterization of three possible architectures for perception and ISL (implicit statistical learning)

formation (e.g., identification of multiple objects as a single triplet) might only be a modification from the usual perceptual process. Like both the modality-specific and standard architectures, the embodied view is able to accommodate the modality- and domain-general nature of ISL. An embodied architecture provides a clear way to accommodate the modality-specific learning effects based upon perceptual conditions such as perceptual grouping: Because learning is embodied in perceptual processing, perceptual biases that are seen in perceptual tasks will robustly transfer to the process of acquiring knowledge through exposure to statistical regularities. Unlike a modality-specific architecture, an embodied mechanism is more parsimonious; it may also better accommodate cross-modal effects in ISL (e.g., learning based on cross-modal statistical regularities could be embodied in multisensory perceptual systems).

It is important to note that there are computational models of learning that are influenced by the relative timing of sensory input. One is *spike-timing-dependent plasticity*, an experimentally observed form of long-term neural plasticity that depends on the relative timing of pre- and post-synaptic action potentials in a process that is computationally similar to Hebbian learning (Song, Miller, & Abbott, 2000). In this learning mechanism, it is possible that the window of integration for pre- and post-synaptic activity could be modified across auditory and visual modalities to affect the optimal timing of learned input. Another computational learning mechanism that is affected by timing of sensory events is temporal difference learning, developed to model the reinforcement or reward learning behaviors. Different models have been proposed to modify the effect of timing on learning (e.g. Daw, Courville, & Touretzky, 2002, 2006). Thus, it is possible the implement different variants of temporal difference reinforcement learning models with different learning outcomes based on

timing of sensory input. In this way, a single learning mechanism, spike-timing-dependent plasticity or temporal-difference reinforcement learning, could be model different optimal timing for learning as we observe across auditory and visual modalities in the current study. This proposal is not entirely capture by any one of the three candidate mechanisms presented in Figure 3.6 but could considered as a combination of an Embodied and Standard Views of statistical learning mechanisms. If the different timing parameters of the learning mechanism would *set* within each modality, the bulk of the learning mechanism could be shared between perceptual modalities and would be similar to the Standard View where there are shared learning resources across modalities. However, the setting of the parameter is an important aspect of the learning mechanism and thus, there is some degree of embodiment of learning, as is captured by the Embodied Learning Mechanism.

In sum, the aim of the current paper is not to conclusively support one type of ISL mechanism but to further elucidate the importance of perceptual processing in learning from statistical regularities. We have highlighted the ways in which these three types of architecture could accommodate modality-specific patterns of learning as reported in the current paper while maintaining domain-general learning abilities. However, we assert that both the modality-specific and the embodied view more readily support differential learning across the visual and auditory modalities. Considering established computational learning mechanisms, it is possible that different aspects of these three types of learning mechanisms are necessary to fully capture the pattern of learning seen across modalities.

A further important question is to understand the origins of these modality-

specific learning differences. For instance, are the perceptual constraints observed here true for all categories of sound stimuli or just for speech sounds? Similarly, is it possible that auditory ISL is more temporally tuned because of language-specific experience? Alternatively, is hearing temporally biased due to generic sensory and/or perceptual factors apart from experience with spoken language?

Our understanding of ISL as an important mechanism in cognition and development has progressed from it being characterized as language-specific, to domain-general and abstract, to current thinking emphasizing the effects of perceptual, attentional, and modality-specific constraints. The evidence appears to disconfirm the idealized conception of ISL as a single, undifferentiated mechanism that operates apart from other perceptual and cognitive constraints. By recognizing and further discovering the complexities governing and affecting the operation of this ubiquitous learning mechanism, we may better understand fundamental processes of language, development, and cognition.

CHAPTER 4

HOW ABSTRACT IS STATISTICAL LEARNING? COMPARING LEARNING ACROSS VISUAL AND AUDITORY PERCEPTUAL MODALITIES IN INFANCY

The empirical content of this chapter was developed in collaboration with Drs. Michael Goldstein, Jen Schwade, & Morten Christiansen and Jennifer Misyak, see Emberson, Misyak, Schwade, Christiansen, and Goldstein (in preparation).

This chapter examines visual- and auditory-SL aged 8-10 months in order to evaluate: 1) whether SL has the same developmental trajectory across perceptual modalities; 2) how learning outcomes compare across vision and audition and to previous research with adult learners. Learning was evidenced in both conditions with opposite directions of preference: Infants in the auditory condition displayed a novelty preference, while infants in the visual condition showed a familiarity preference. Next, we found equal magnitudes of learning across modalities. Finally, we found evidence of different developmental trajectories across modalities. The magnitude of auditory-SL changed with age. Using a median split, older infants exhibited a novelty preference while younger infants showed no systematic preference. A comparable group of younger infants showed visual-SL. Thus, we find evidence for earlier availability of SL in the visual modality. This is the opposite pattern from adults where auditory-SL tends to be more robust than visual-SL, suggesting that SL development could be supported by perceptual changes. Overall, these findings suggest that SL is not amodal but permeable to perception.

4.1 Introduction

The human brain has the remarkable ability to use knowledge of underlying structure to shape perception. While this perceptual ability could be the result of innate knowledge, there is increasing evidence that perception can be shaped through experience. Environmental structure is available to the learner through statistical regularities in sensory input; statistical learning is the process of detecting these regularities after incidental exposure. With demonstrations that infants can engage in statistical learning after only a brief exposure (Saffran et al., 1996; neonates in the auditory modality: Teinonen, Fellman, Näätänen, Alku, & Huotilainen, 2009; 2-month-olds in the visual modality; Kirkham et al., 2002), the last decade of research in developmental psychology has focused on the role of statistics in shaping perception and cognition. Specifically, it is believed that prior experience with statistical regularities can support development across a number of domains such as language learning (Romberg & Saffran, 2010), object and scene perception (Fiser & Aslin, 2002), and music perception (McMullen & Saffran, 2004). There is also evidence that infants as young as 8 months make inferences based on statistical information (Xu & Garcia, 2008) which can in turn support development in additional cognitive domains such as causal reasoning (Gopnik & Tenenbaum, 2007).

Despite the importance of statistical learning to perceptual and cognitive development, very little is known about the nature and the development of the mechanisms underlying statistical learning (SL): Is SL developmentally invariant or is there a discernible developmental trajectory? Are the underlying mechanisms amodal and abstract, and thus sensitive only to the statistical regularities present in sensory input, or is learning modulated by non-statistical perceptual

differences (e.g. whether the statistics are presented as part of auditory vs. visual stimuli)? For clarity, the paper will refer to these factors as "statistical information" and "perceptual information," respectively.

Recent research in adult learners has demonstrated that SL can differ across auditory and visual perceptual modalities even if the statistical information is held constant (e.g., Conway & Christiansen, 2005; Saffran, 2002). Moreover, a number of studies have suggested that auditory SL is superior to visual SL in adults (Conway & Christiansen, 2005; Emberson et al., 2011; Robinson & Sloutsky, 2007).

While this work suggests that SL is not an amodal, abstract learning mechanism, it is unknown how infant SL compares across perceptual modalities. Separate studies have found that infants can learn using both visual and auditory stimuli (e.g., Fiser & Aslin, 2002; Kirkham et al., 2002; Saffran et al., 1999). However, studies examining auditory and visual SL employ different methodologies making it difficult to compare learning outcomes and, to date, no single study has used the same methodology to examine differences of visual and auditory SL with infant learners.

If SL is also sensitive to perceptual information, in addition to statistical information, in infancy, it is unknown whether or how the interaction of perception and SL develops. Previous studies have suggested that SL abilities are invariant across developmental time. Saffran et al. (1997)compared auditory SL between adults and children and Kirkham et al. (2002) examined visual SL in 2 to 8-months-old infants; both studies found no difference in learning outcomes across age groups. However, there has been some recent evidence suggesting that SL abilities can develop. (Thiessen et al., 2005) suggested that increased

age (as well as increased exposure) can make a learning task easier in infancy. Arciuli and Simpson (2011) directly examined visual SL in children 5- to 12-years-of-age and found that their ability to differentiate familiar and novel items at test increased across childhood. Recent work by McNealy, Mazziotta, and Dapretto (in press) found differences in the neural activation associated with auditory SL between 6 and 10-year-old learners, suggesting changes in auditory SL across childhood. Thus, while studies suggesting developmental invariance of SL are compelling, they are sparse (visual SL in 2-8 months: Kirkham et al., 2002; auditory SL in children and adults: Saffran et al., 1997) and recent evidence has emerged suggesting SL abilities may increase across developmental time.

If SL is developmentally invariant, one would predict the same pattern of modality-specificity in infancy as is seen in adulthood (i.e. better auditory than visual SL in the canonical SL paradigm, in which learners use transitional probabilities to segment items in a stream as presented in Saffran et al., 1996). However, if SL is modulated by the type of perceptual information and develops over time, one might expect developmental trajectories to differ across perceptual modalities. Such differences in development could be supported by changes in perceptual processing across visual and auditory sensory modalities and would likely not be the result of changes in an abstract learning mechanism.

The current study compared auditory and visual SL in infants 8- to 10-months-old to examine 1) how learning outcomes compare across vision and audition and to previous research with adult learners, and 2) whether SL has the same developmental trajectory across perceptual modalities. To gain a clear view of differences in learning outcomes, we strove to equate learning condi-

tions across modalities. First, we employed virtually identical methodologies for visual and auditory learning paradigms. Previous visual SL studies have typically employed infant-controlled habituation (Fiser & Aslin, 2002; Kirkham et al., 2002) while auditory SL studies have employed fixed familiarization to sounds (e.g., Saffran et al., 1996). Some studies have presented fixed familiarization with infants engaging in other tasks such as watching a cartoon or playing with their caregivers (e.g., Graf Estes et al., 2007). This difference in methodology stems from basic differences across vision and audition (input of visual information requires the eyes to be pointed at the relevant stimuli while one can hear ambient auditory stimuli without any overt orienting) and also the assumption that auditory SL can occur without selective or focused attention (Saffran et al., 1997). To equate methodologies across modalities and to be able to compare amount of exposure to statistical information, we employed an infant-controlled habituation paradigm in both visual and auditory conditions.

Second, we aimed to equate the*type* of stimuli: previous visual SL studies have employed geometric shapes (Fiser & Aslin, 2002; Kirkham et al., 2002) while auditory SL studies have typically used speech sounds (Saffran et al., 1996; however see studies with tones, Creel, Newport, & Aslin, 2004; Saffran et al., 1999). Later in their first year of life, infants have had considerable exposure to speech sounds making them more familiar than geometric shapes; moreover, speech sounds are more perceptually complex and infants are beginning to develop specialized processing for speech (e.g., Kuhl et al., 1992). Thus, we compared SL using speech sounds (auditory condition) and faces (visual condition) as both of these stimuli are familiar, perceptually-complex, and the subject to specialized processing for infants in our age range (Nelson, 2010; Pascalis et al., 2005).

Finally, visual and auditory SL studies with infants have often employed different rates of stimulus presentation. Visual SL stimuli have often been presented at a slower rate than auditory SL stimuli (e.g. visual SL: 1 stimulus/second; Kirkham et al., 2002; auditory SL: 4-5 stimuli/second; Saffran et al., 1996; Pelucchi, Hay, & Saffran, 2009). Increased rates of presentation have been shown to decrease visual SL in children (Arciuli & Simpson, 2011) and adults (Turk-Browne et al., 2005). Research in adults has suggested the opposite effect with auditory SL: decreased learning at slower rates of presentation (Emberson et al., 2011). These results suggest that different modalities require different rates of presentation to support SL. The relationship between rate, SL, and perceptual modality has not been directly tested in infant learners; however, there is circumstantial evidence that such a relationship exists. Specifically, prior studies finding evidence for visual SL tend to have slower rates of presentation than studies finding evidence for auditory SL, as noted above. Thus, we choose rates which balanced the constraints of achieving similarity across modalities and with the rate of presentation likely required by that perceptual system (visual rate of presentation: 1 stimulus/second the same as Kirkham et al., 2002; auditory rate of presentation: 2 stimuli/second similar to Thiessen et al., 2005).

4.2 Methods

4.2.1 Participants

Fifty-nine infants between the ages of 8- and 10-months were recruited for the current study. Twenty-six infants were recruited for the Auditory condition and

33 infants were recruited for the Visual condition. Within each condition, infants were counterbalanced for sex, two sets of bigrams orders, and test trial order. Nineteen additional infants were tested but excluded because of excessive fussiness (11: auditory; 6 visual), preterm birth (1:auditory) or parental interference (1: visual). Infants were recruited at experiment locations in Ithaca (n = 50) and Syracuse (14 in the visual condition only).

Visual examination of infants looking to the test trials revealed significant outliers in each Perceptual Condition (e.g. one infant looked longer at the Novel test trials for an average of 48 seconds; the mean difference was close to 2 seconds). Thus, infants with looking scores during test (Difference Score = Mean Looking to Novel Familiar Test Trial) of greater than two standard deviations (*SD*) from the mean were excluded from further analyses.[1] Means and *SDs* were calculated separately for each Perceptual Condition and habituation group. This resulted in the exclusion of five additional infants (Habituated infants: 2 auditory; 1 visual; Non-Habituated infants: 1 auditory, 1 visual).

Of the 59 infants included in the subsequent analyses, 32 were female, 26 were assigned to the Auditory condition, and the mean age was 9.08 months (*SD* = 0.64, range = 7.9 to 10.0 months).

4.2.2 Stimuli and Statistical Sequences

The current study employed equivalent sets of visual and auditory stimuli. Six smiling, Caucasian, female faces were selected from the NimStim database (Fig-

[1]This criterion for excluding outliers has been previously applied in a number of related studies on infant learning (e.g. Haan & Nelson, 1997; Maye, Weiss, & Aslin, 2008) and SL (e.g., Conway, Bauernschmidt, Huang, & Pisoni, 2010).

Figure 4.1: All six face stimuli used in the current experiment in a sample bigram order.

ure 4.1; Tottenham et al., 2009). Faces were presented individually at a rate of 1 stimulus / second (1 second SOA). Six monosyllabic nonwords (e.g., vot) were recorded with equal lexical stress and flat prosody (adult-directed or AD speech) from a female native English speaker. The length of each utterance was edited to have a uniform duration of 600 ms and stimuli were presented at a rate of 2 stimuli/second (500ms SOA). Nonwords were played acoustically and accompanied by the projected image of a checkerboard (4 x 4 black and white, with gray surround) to direct the infants attentional focus. Both face stimuli and the checkerboard used for the auditory condition subtended 14.6° of visual

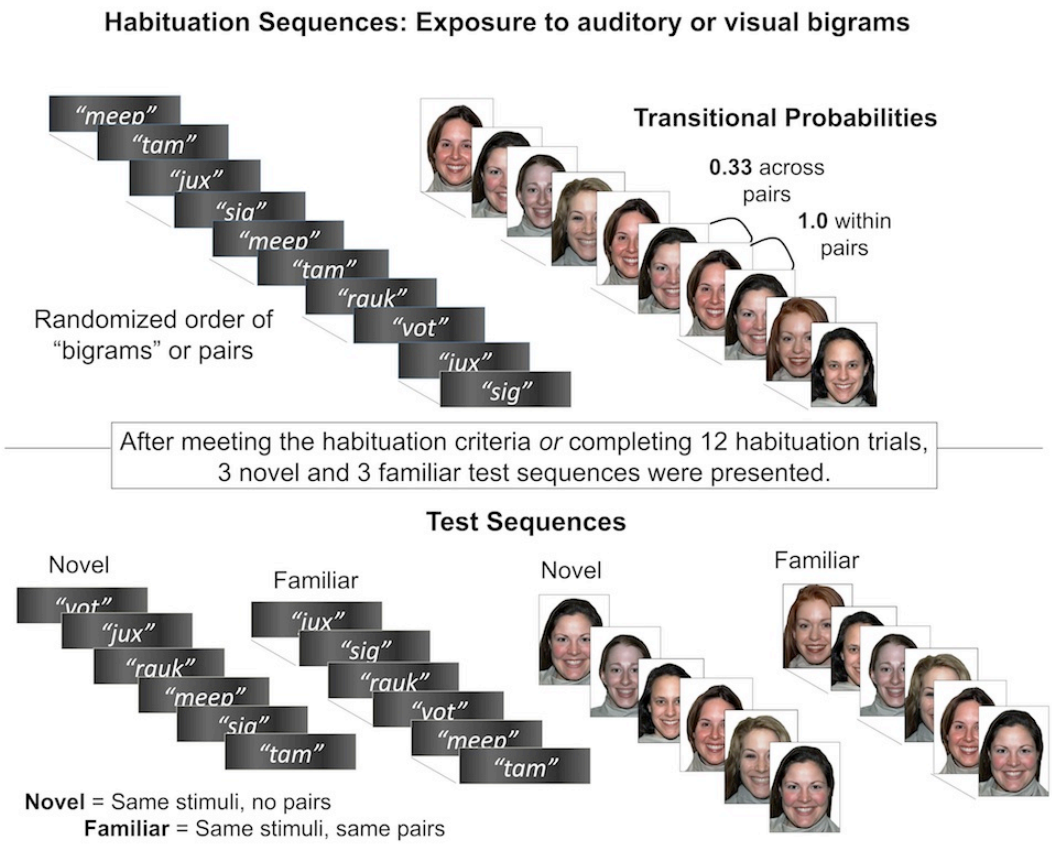


Figure 4.2: Depiction of habituation, including transitional probabilities, and test sequences (familiar and novel) for the visual and the auditory perceptual conditions (left and right). Each stimulus was presented individually with the order to stimuli presented along the diagonal in the figure.

angle (Kirkham et al., 2002). Auditory stimuli were presented at 58 dB.

For each condition (visual or auditory), the six stimuli (faces or nonwords) were grouped into two mutually exclusive sets of "bigrams" or pairs (see Figure 4.1 for sample face bigram set). Each infant was only exposed to one bigram set. This procedure allowed us to examine whether learning outcomes were affected by the particular grouping of the stimuli.

Habituation sequences were constructed by concatenating three bigrams of

a given set in random order with no more than four consecutive presentations of a single bigram (Figure 4.2). All bigrams were presented with equal frequency. The only cue to bigram structure was the statistical information in the stream: Both co-occurrence frequencies and transitional probabilities could support bigram segmentation (Aslin et al., 1998). Transitional probabilities between successive stimuli were higher within pairs (1.0) than between pairs (0.33). These transitional probabilities have been shown to be learnable by infants in the tested age range (Kirkham et al., 2002; Saffran et al., 1996). Twelve different habituation sequences were constructed for each bigram order for each condition.

There were two types of test trial sequences: familiar and novel. The only difference between test trial types was the statistical information of the presentation streams. This ensured that any looking time differences were due to infants' sensitivity to the statistical structure of the sequences. Familiar trials were constructed using identical methods as the habituation sequences. Novel trials were constructed by using a random order of all stimuli (i.e. no consistent bigram pairings and no reliable statistical information), with the constraint that a single stimulus could not be presented consecutively more than four times. Three novel and three familiar test trials were constructed for each bigram ordering and for each condition.

Individual stimuli (and bigrams) were presented at equal frequencies for all sequence types. Both habituation and test trial sequences were 60 seconds long, thus visual sequences contained 60 stimuli while auditory sequences contained 120 stimuli.

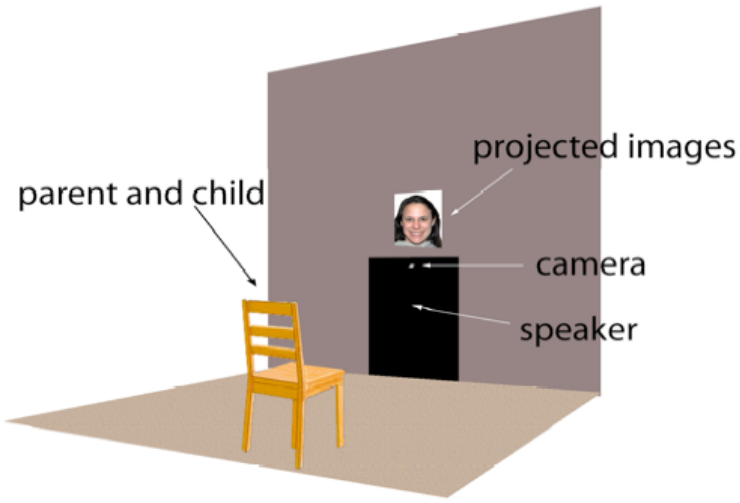


Figure 4.3: Depiction of the experimental set up. Infants sat on their caregivers lap in the chair while stimuli were presented. All visual stimuli (a checkerboard during the auditory condition, faces in the visual condition, and the attention getter, used in both conditions) were presented centrally. Auditory stimuli (speech tokens or the sound for the attention getter) were presented from a speaker placed in front of the infants and below the visual stimuli.

4.2.3 Procedure

Infants were tested individually, seated in a caregiver's lap in a darkened room. Their caregivers were instructed to keep their infants on their laps facing forward, if possible, but not to interfere with their looking or behavior. To further prevent any unconscious cueing or interference, each caregiver wore sound attenuating headphones with music playing and either a hat or a visor that prevented visual access to the stimuli. Images were presented centrally onto the wall or a screen in front of the infant. Sounds were presented centrally directly below the visual stimuli (Figure 4.3).

Stimulus presentation was controlled by Habit 2000 (Cohen, Atkinson, & Chaput, 2000) operating on a Macintosh computer running OS 9. Directly be-

low the projected images, but above the speaker, a camera was mounted to record infant eye gaze (Figure 4.3). An observer in a different room, blind to the sequences and trial types, recorded looks towards and away from the visual stimuli (i.e. faces for the visual condition and the checkerboard for the auditory condition).

Infants were presented with an attention-getting animation of a rotating and looming blue and white disc with sound in between trials. This attention-getter would play until the infants looked centrally at which point a sequence was presented. If the infant did not look at the beginning of the sequence for at least two seconds, this was not counted as a completed trial; the attention-getter played again and once the infant looked centrally, the same sequence was repeated. If the infant looked for two seconds or longer, the sequence would play until infants looked away for two consecutive seconds or the sequence ended (60 seconds; Kirkham et al., 2002; Saffran et al., 1996).

Habituation sequences were presented in random order until infants either reached the habituation criterion or all 12 habituation sequences had been presented. The habituation criterion was defined as a decline of looking time by more than 50% for four consecutive trials, using a sliding window, compared to the first four habituation trials. Infants were then presented with six test trials in alternating order by test trial type (familiar and novel) with the order of alternation (i.e. novel first or familiar first) counterbalanced across infants (Kirkham et al., 2002).

4.3 Results

4.3.1 Habituation and Viewing Time

Visual Condition.

Eighteen infants met the habituation criterion, Habituated infants, with a mean of 8.1 habituation trials (SD = 2.3) and a viewing time during habituation of 105.4 seconds (SD = 68.3). The 15 infants who did not reach habituation criterion, Non-Habituated infants, had a mean viewing time of 130.2 seconds (SD = 48.6). There was no difference in viewing time during habituation between these groups, $t(31) = 1.18$ $p = 0.25$, $d = -0.42$.

Auditory Condition.

Seventeen infants met the habituation criterion with a mean of 8.7 trials (SD = 2.2) and viewing time of 75.0 seconds (SD = 37.9). The nine infants who did not reach habituation criterion had a mean viewing time of 118.36 seconds (SD = 39.8). Habituated infants looked for significantly less time than Non-Habituated infants, $t(24) = -2.72$, $p = 0.012$, $d = -1.12$.

Considering differences in habituation across perceptual modalities, we found no significant difference in viewing time for Non-Habituated infants across perceptual modalities, $t(22) = 0.624$, $p = 0.54$, $d = 0.27$, or for Habituated infants, $t(32) = -1.61$, $p = 0.12$, $d = -0.55$. We also investigated whether there is a difference in the distribution of Habituated vs. Non-Habituated infants across perceptual modalities (Chi-squared test with factors of Habituation

group and Perceptual Modality). We failed to reject the null hypothesis, χ^2 (1, n = 59) = 0.708, p = 0.40, supporting the view that there is no difference in the distribution of habituated to non-habituated infants across perceptual modalities. For subsequent analyses, unless noted otherwise, infants were compared within Habituation group.

4.3.2 Evidence of Learning

In the current task and consistent with previous literature (e.g., Jusczyk & Aslin, 1995; Saffran & Thiessen, 2003; Thiessen et al., 2005), we define evidence of learning as a significant difference in looking between test trial types (Novel vs. Familiar).

Visual Condition.

We investigated whether looking to test trial types (Novel vs. Familiar) was uniform across Habituation groups. A mixed ANOVA (within subject factor: test trial type, between subject factor: Habituation group) revealed no main effects of either test trial or habituation group, $ps > 0.5$, but a significant interaction of test trial type and habituation group, $F(1, 31) = 12.86$, $p = 0.01$, $\eta^2 = 0.293$, suggesting that infants do not look uniformly to test trial types and this difference is mediated by whether they reached the habituation criterion. Following up this interaction, we investigated the pattern of looking for separately across habituation groups.

Habituated infants in the Visual condition show evidence of learning; they

looked at the Novel test trials for a mean of 6.11 seconds (SD = 2.6) and to the Familiar trials for 8.24 seconds (SD = 3.8; Figure 4.4). Paired t-tests revealed a significant difference in looking between trial types ($t(17)$ = -3.46, p = 0.003, d = -0.896). This difference was confirmed in non-parametric paired tests (15 of the 18 infants showed a bias toward looking to the familiar trials, Wilcoxon signed-rank test, Z = -2.72, p = 0.006). Thus, habituated infants in the visual condition showed a significant familiarity preference.

While Non-Habituated infants looked longer at Novel test trials (Novel: M = 8.8 seconds, SD = 4.4; Familiar: M = 6.85, SD = 3.92), this difference only reached marginal significance in a paired parametric test ($t(14)$ = 1.94, p = 0.073, d = 0.51). There was no difference across trial types in a non-parametric test (9 of 15 infants showed bias towards the Novel trials, Wilcoxon signed-rank test, Z = -1.1, p = 0.26). Therefore, we found robust evidence of learning in Habituated infants only in the Visual condition: Habituated infants exhibit a significant familiarity preference during test.

Auditory Condition.

A mixed ANOVA (within subject: test trial type; between subject: habituation group) again revealed a significant interaction between habituation and looking across trial types, $F(1, 24)$ = 11.03, p = 0.003 , η^2 = 0.31. We also find a main effect of habituation group, $F(1,24)$ = 4.88, p = 0.037, η^2 = 0.169, which is supported by significantly longer looking by Non-Habituated infants overall during test, $t(24)$ = -2.72, p = 0.012, d = -1.11. There is no effect of test trial type, $F(1, 25)$ = 0.083, p = 0.78, η^2 = 0.002. To follow up the interaction of test trial type and habituation group, we investigated looking across trial types within Habituated

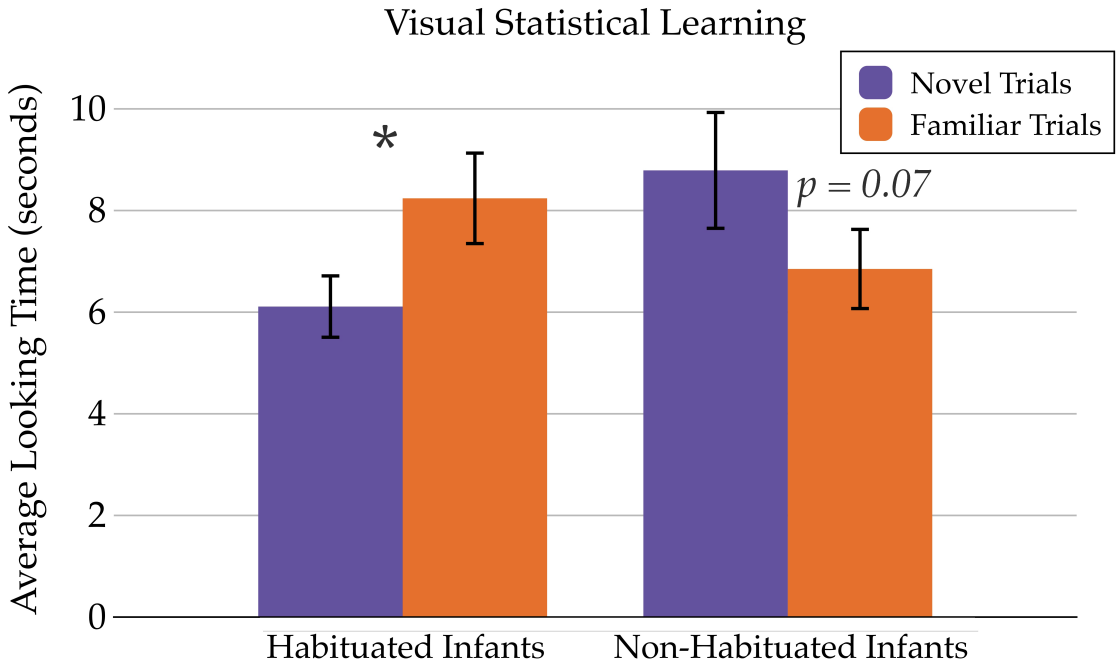


Figure 4.4: Learning outcomes for the visual statistical learning condition.

and Non-Habituated infants.

We also found evidence for learning for Habituated infants in the Auditory condition: they looked at the Novel trials for a mean of 5.16 seconds (SD = 1.6) and to the Familiar test trials for 3.92 seconds (SD = 1.4; Figure 4.5). Paired t-tests revealed a significant difference in looking time across these trial types, $t(17)$ = 3.126, p = 0.006, d = 0.74. Non-parametric tests confirm this difference, (13 out of 18 infants showed bias towards the Novel trials, Wilcoxon signed-rank test, Z = -2.68, p = 0.007). Thus, Habituated infants in the auditory modality also showed evidence of learning with a significant novelty preference.

Examining Non-Habituated infants, we again found a marginally significant difference in looking across test trials with an opposite direction of preference to the Habituated infants: they looked at the Novel trials for a mean of 5.88 seconds (SD = 3.4) and at the Familiar trials for 6.93 seconds (SD = 3.0). A paired t-test

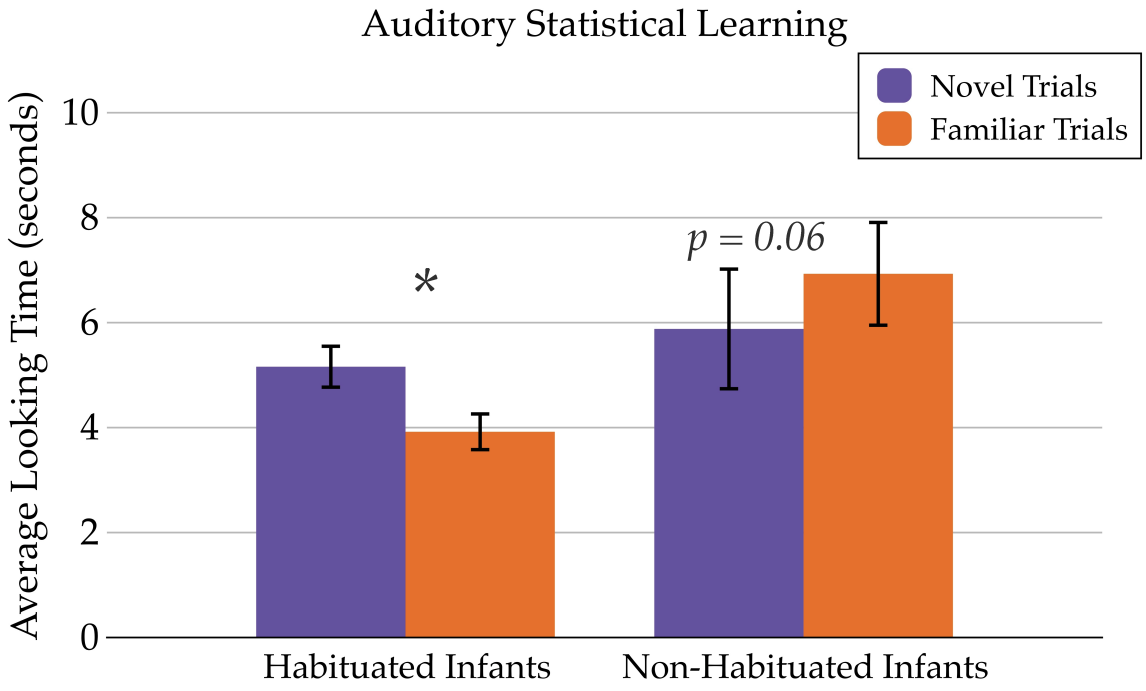


Figure 4.5: Learning outcomes for the auditory statistical learning condition.

revealed a marginally significant difference between the two trial types, $t(8)$ = -2.20, p = 0.059, d = -0.33. Non-parametric tests also indicated a marginally significant difference (8 out of 9 infants showed a Familiarity preference, Wilcoxon signed-rank test, Z = -1.96, p = 0.051). Thus, we find evidence for a significant Novelty preference for Habituated infants in the Auditory condition and some indication that Non-Habituated infants exhibit a familiarity preference.

4.3.3 Comparing Learning Across Perceptual Modalities

The goal of the current study is to compare SL across the visual and auditory perceptual modalities in infancy. First, we aimed to directly investigate whether learning in different perceptual modalities exhibits a similar pattern of preference at test. To this end, mean looking times of Habituated infants were sub-

mitted to a mixed ANOVA (within subjects factor: novel vs. familiar test trials; between subjects factor of perceptual modality: visual vs. auditory). This analysis revealed a main effect of Perceptual Modality, $F(1, 33) = 11.68$, $p = 0.002$, $\eta^2 = 0.26$, and an interaction of Perceptual Modality and Test Trial Type, $F(1, 33) = 19.71$, $p < 0.001$, $\eta^2 = 0.37$ (Figure 4.6). Follow up independent sample t-tests revealed that the main effect of modality is driven by longer looking, in general, in the visual condition (M = 14.4 sec, SD = 5.9) than infants hearing auditory sequences (M = 9.08 sec, SD = 2.4), equal variances not assumed, $t(22.8) = -3.49$, $p = 0.002$, $d = -1.18$. This effect could be the result of the slower rate of presentation of the visual sequences. However, we only find marginally longer looking at visual sequences during habituation. The interaction of perceptual modality and test trial type is likely driven by the opposite preferences at test across perceptual modalities, with infants in the Visual modality showing a significant familiarity preference and infants in the Auditory modality showing a significant novelty preference. While this result suggests that infants do not respond uniformly to statistical information across perceptual modalities, as evidenced by the interaction of modality and test trial type, the significantly longer looking at test by infants in the visual condition makes it difficult to use the raw looking data to directly compare learning outcomes across modalities.

In order to control for the generally longer looking in the Visual modality, we examined looking to Novel and Familiar test trials calculated as a proportion of the total mean looking to both test trial types (Figure 4.7). Thus, any difference in the proportion of looking to one trial type or another will not result from general differences in the magnitude of looking across perceptual modalities. To compare proportionate looking to Novel vs. Familiar test trials across modalities, we again submitted the data for Habituated infants to a mixed ANOVA

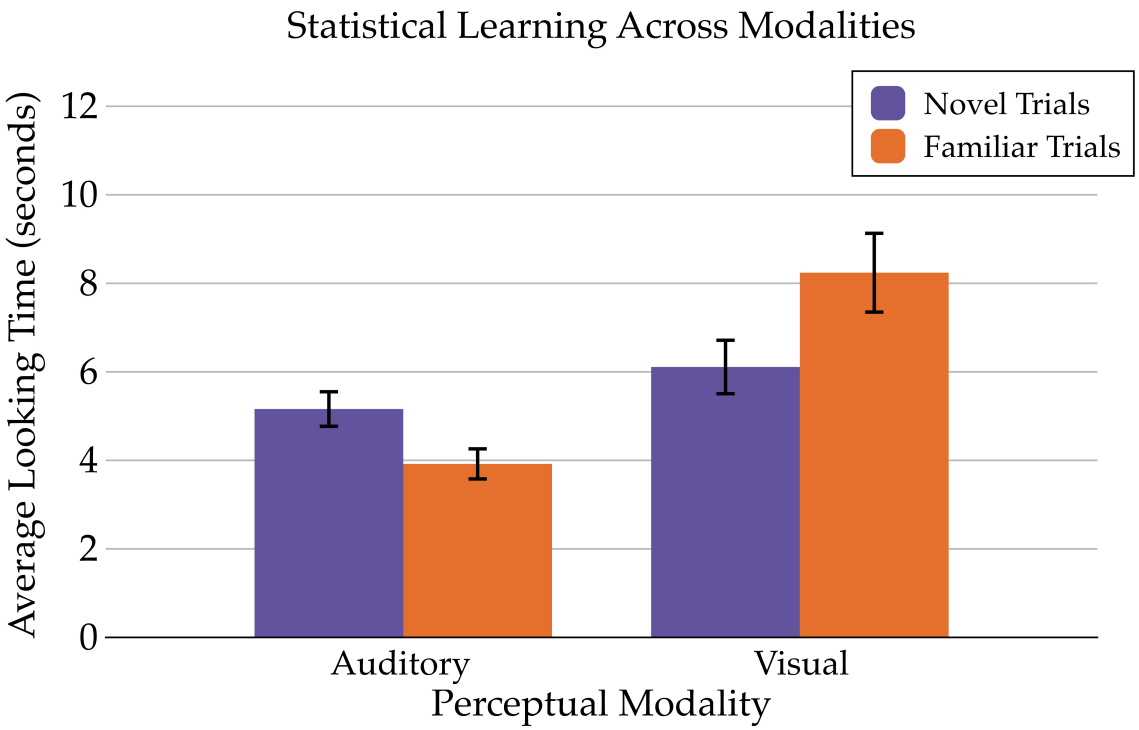


Figure 4.6: Looking to Novel and Familiar test trials for Habituated infants perceiving visual and auditory sequences.

(within subject factor: trial type; between subjects: perceptual modality). We found no significant main effect of Perceptual Modality confirming that we have controlled for the main effect of modality on looking time. However, we still found a significant interaction of test trial type and perceptual modality, $F(1, 33) = 14.70$, $p = 0.001$, $\eta^2 = 0.31$ Follow up t-tests again confirmed that these results are due to opposite, but significant, directions of preference at test (Visual: $t(17) = -2.70$, $p = 0.015$, $d = 1.28$; Auditory: $t(17) = 2.84$, $p = 0.011$, $d = 1.32$). In sum, we established that there are differences in preference at test across perceptual modalities and differences at test do not stem from general differences in looking at test across perceptual modalities.

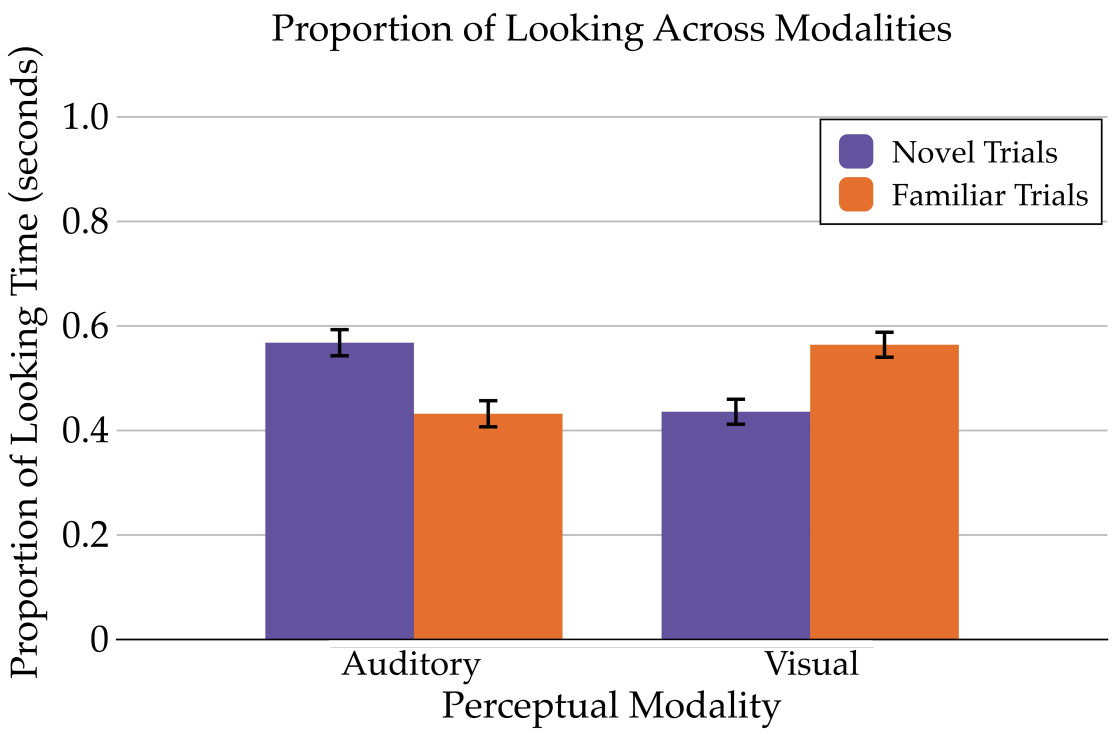


Figure 4.7: Proportion of looking time to novel and familiar test trials across modalities. This analysis controlled for the general differences in looking time between perceptual modalities and still revealed a significant modality by test trial type interaction.

4.3.4 Magnitude of Learning Across Perceptual Modalities

Next, we aimed to consider whether the magnitude of learning is different when infants receive auditory or visual statistical information. While SL studies typically discuss learning outcomes as a binary measure of learnability (yes vs. no), the magnitude of learning, a continuous measure, could be considered as the difference in looking between novel and familiar test trials. The assumption of this measure is that better learning would result in greater differences in looking across novel and familiar test trials and vice versa. Visual inspection of Figure 4.7 reveals a striking symmetry in difference in the distribution of looking between test trial types. However, as established above, infants in the auditory

and visual modalities exhibit different directions of preference at test. Thus, we cannot include directionality in a measure of the magnitude of learning. To this end, we calculated a Difference Score (mean duration of looking to the Novel - Familiar test trials) and a Proportionate Difference Score (mean proportion of looking to the Novel -Familiar test trials) for each infant. We then took the absolute value of these measures and used independent sample t-tests to compare differences in looking across perceptual modalities in Habituated infants. We found a marginally significant difference in the magnitude of learning for the Absolute Difference Scores (visual: M = 2.79 seconds, SD = 1.9; auditory: M = 1.84, SD = 1.1 seconds; $t(33)$ = -1.81, p = 0.08, d = 0.61). However, this measure does not control for generally longer looking to visual test trials. We do not find evidence for a difference in magnitude of learning in the Absolute Proportionate Difference Scores (visual: M = 0.201, SD = 0.12; auditory: M = 0.21, SD = 0.12; $t(33)$ = .221, p = 0.83, d = 0.074) suggesting that the generally longer looking to Visual test trials could drive the marginal difference in the Absolute Difference Score. Overall, these analyses suggest that there is no difference in the magnitude of learning across perceptual modalities when the direction of preference is not taken into account and when generally longer looking to visual sequences is accounted for.

4.3.5 Considering Learning Outcomes in Relation to Rate of Presentation

One difficulty in comparing the magnitude of learning is the difference in rate of presentation employed for visual (1 token/second) and auditory stimuli (2 to-

kens/second). The rate of visual SL is consistent with previous studies in both infants (Kirkham et al., 2002) and adults (Turk-Browne et al., 2005). An equivalent rate of presentation for the auditory SL would be much slower than rates previously reported in the literature (see Table 4.1). Moreover, recent results with adult learners have suggested that auditory SL decreases with decreased rate of presentation while the opposite is true for visual SL (Emberson et al., 2011).

We sought to confirm that, in order to elicit SL, auditory and visual modalities require presentations at different rates. To this end, we conducted a control study using identical methods that presented the auditory SL stimuli at a rate of 1 token/second, the equivalent rate to the visual condition. Fourteen infants were included in the final sample with a mean age = 9.3 months (SD = 0.53). Habituated infants (n = 9) looked at the Novel test trials for M = 6.1 seconds, SD = 2.8 and at the Familiar trials for M = 5.9 seconds, SD = 1.1. Non-habituated infants looked at the Novel test trials for M = 4.6 seconds, SD = 1.9, and at the Familiar for M = 4.4 seconds, SD = 1.1. There was no evidence of learning for habituated nor non-habituated infants in either parametric or non-parametric tests (Habituated: 5 of 9 infants exhibited a Novelty preference, Wilcoxon signed rank test, Z = -0.06, $p > 0.9$; $t(8) = 0.2$, $p > 0.8$, d = .14; Non-Habituated: 2 of 5 infants exhibited a Novelty preference, Wilcoxon Signed rank test, Z = -0.4, $p > 0.6$; $t(4) = 0.2$, $p > 0.8$, d = 0.2). Thus, we find that there is no evidence of learning for infants who receive auditory familiarization at the same rate of presentation as the visual familiarization (1 token/second). These results establish the necessity to present auditory tokens at a faster rate of presentation than the visual tokens in the current study.

Next, we considered learning outcomes while controlling for differences in rate of presentation by considering looking times in terms of the number of stimuli perceived rather than amount of time. Because auditory and visual stimuli were presented at different rates (2 stimuli/second and 1 stimulus/second, respectively) in order to elicit SL, infants in the Auditory condition perceived double the number of tokens during the same amount of viewing time as an infant in the Visual condition. In order to consider viewing time during habituation in terms of number of tokens perceived, looking times for infants in the Auditory condition were multiplied by two while the looking times remained the same for the Visual condition.

Considering the number of tokens perceived during habituation, Habituated infants in the Auditory condition perceived 150 tokens (75 bigrams, SD = 76 tokens) while infants in the Visual condition perceived 105 tokens (52.5 bigrams, SD = 68 tokens). The difference in number of tokens perceived across perceptual modalities was marginally significant ($t(33) = 1.83$, $p = 0.076$, $d = 0.62$). Non-Habituated infants viewed significantly more tokens in the auditory than the visual condition (Visual: M = 130 tokens, SD = 48; Auditory: M = 237 tokens, SD = 80, unequal variances not assumed, $t(11.5) = -3.6$, $p = 0.004$, $d = 1.62$).

Next, we submitted number of tokens perceived during test by Habituated infants to a mixed ANOVA with test trial type (novel vs. familiar) as a within subject factor and modality as a between subject factor (Figure 4.8). This analysis revealed a main effect of modality, $F(1, 33) = 4.32$, $p = 0.045$, $\eta^2 = 0.12$. In contrast to the results based on looking times, this main effect is a result of infants hearing more stimuli in the Auditory condition during test. We also find a significant interaction of test trial type and modality $F(1, 33) = 19.3$, $p < 0.001$,

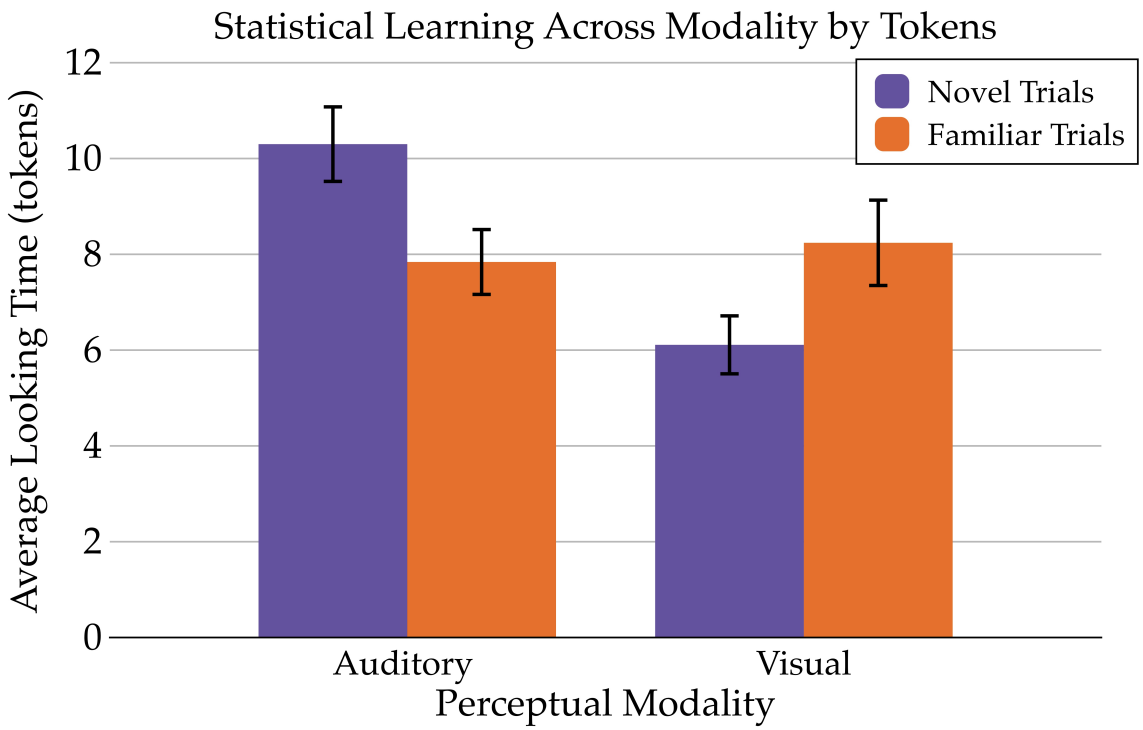


Figure 4.8: Comparisons of looking times at test in terms of number of tokens or stimuli perceived. This analysis is meant to control for differences in rate of presentation.

$\eta^2 = 0.37$. As before, this result is likely driven by differences in the direction of preference across perceptual modalities at test.

Having controlled for the rate of presentation, we return to the question of differences in magnitude of learning across modalities. Comparison of the absolute value of the Difference in Tokens Perceived (mean tokens perceived for the Novel - Familiar test trials) again revealed no difference in the magnitude of learning, $t(33) = 1.29$, $p = 0.21$, $d = 0.43$. Thus, even while accounting for differences in the number of tokens perceived during an equivalent amount of looking time, we find no differences in the magnitude of learning across modalities.

4.3.6 Influence of Viewing Time during Habituation on Learning

Previous research has suggested that longer passive exposure to statistical regularities results in better learning (Gebhart, Aslin, & Newport, 2009; Thiessen et al., 2005). As noted above, Non-Habituated infants viewed habituation stimuli for a longer period of time than Habituated infants and yet we found no consistent evidence of SL in this group. We examined whether viewing time influences SL in Habituated infants by correlating viewing time with Difference scores–mean looking to novel – familiar test trials. While we found no significant influence of viewing time on learning for infants in the Visual condition (n = 18, $p > 0.5$), we did find a significant positive correlation for infants in the Auditory condition (r = 0.49, n = 17, p = 0.046, Figure 4.9). There is no relationship between viewing time and Difference score for Non-Habituated infants nor when pooling infants across Habituation groups. Thus, infants who look longer during habituation tend towards a stronger novelty preference but only when they heard statistical information in auditory sequences.

4.3.7 Influence of Age on Learning

We also examined whether age influenced learning outcomes in Habituated infants. Again, we found no significant correlation between age and Difference Score for infants in the Visual condition (r = 0.35, n = 18, p = 0.16) but there was a marginally significant correlation of age with Difference Score, r = 0.58, n = 17, p = 0.015, for infants in the Auditory condition with older infants exhibiting a stronger Novelty preference (Figure 4.10). Examining infants in each modality

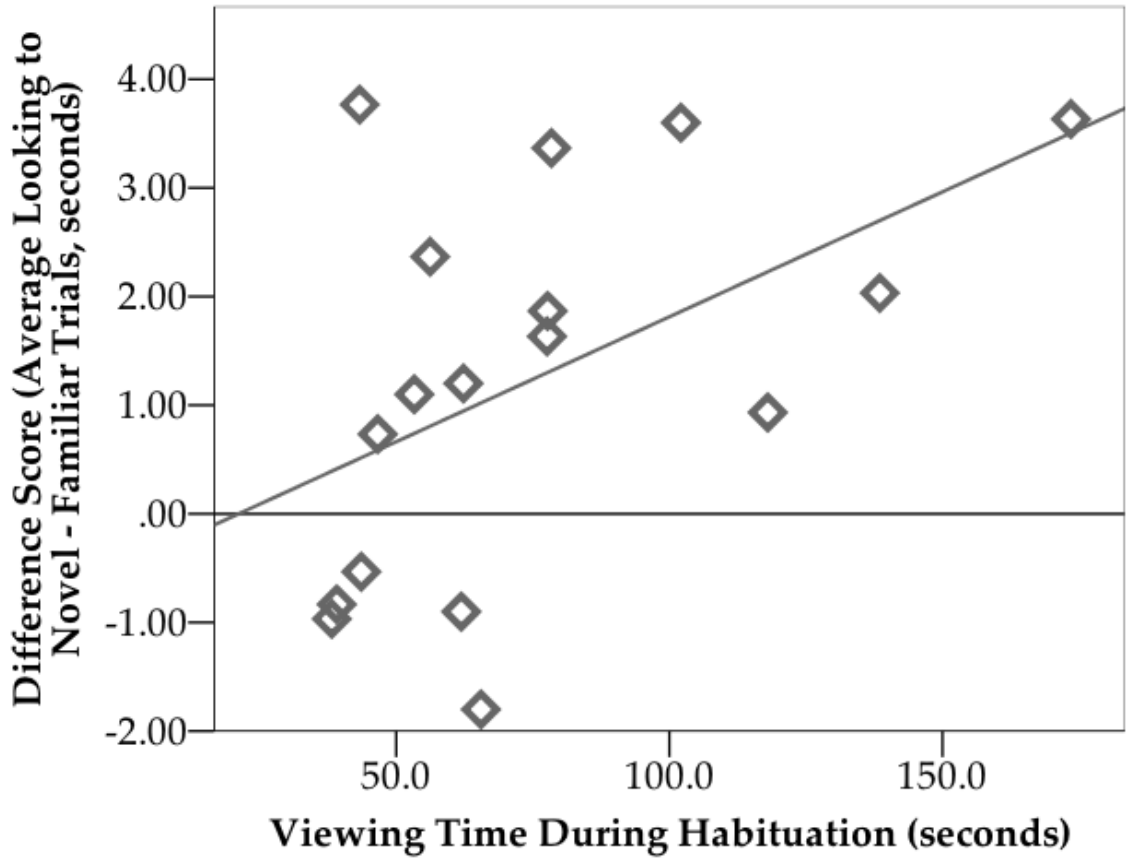


Figure 4.9: Habituated infants in the auditory condition show a significant correlation between viewing time during habituation (seconds) and Difference Score with infants with longer looking times showing a stronger novelty preference.

regardless of habituation status, we find a significant correlation of age and Difference score in the Auditory modality, $r = 0.56$, $n = 26$, $p = 0.003$, but again not in the Visual modality, $r = 0.03$, $p = 0.68$.

Given that infants that have viewing time and age are both predictors of learning outcomes in the Auditory condition, we tested whether there is a correlation between age and viewing time. We find no significant correlation ($r = .30$, $n = 17$, $p = 0.24$). Thus, we find that age and viewing time both positively correlate with Difference Score, however, the lack of significant correlation be-

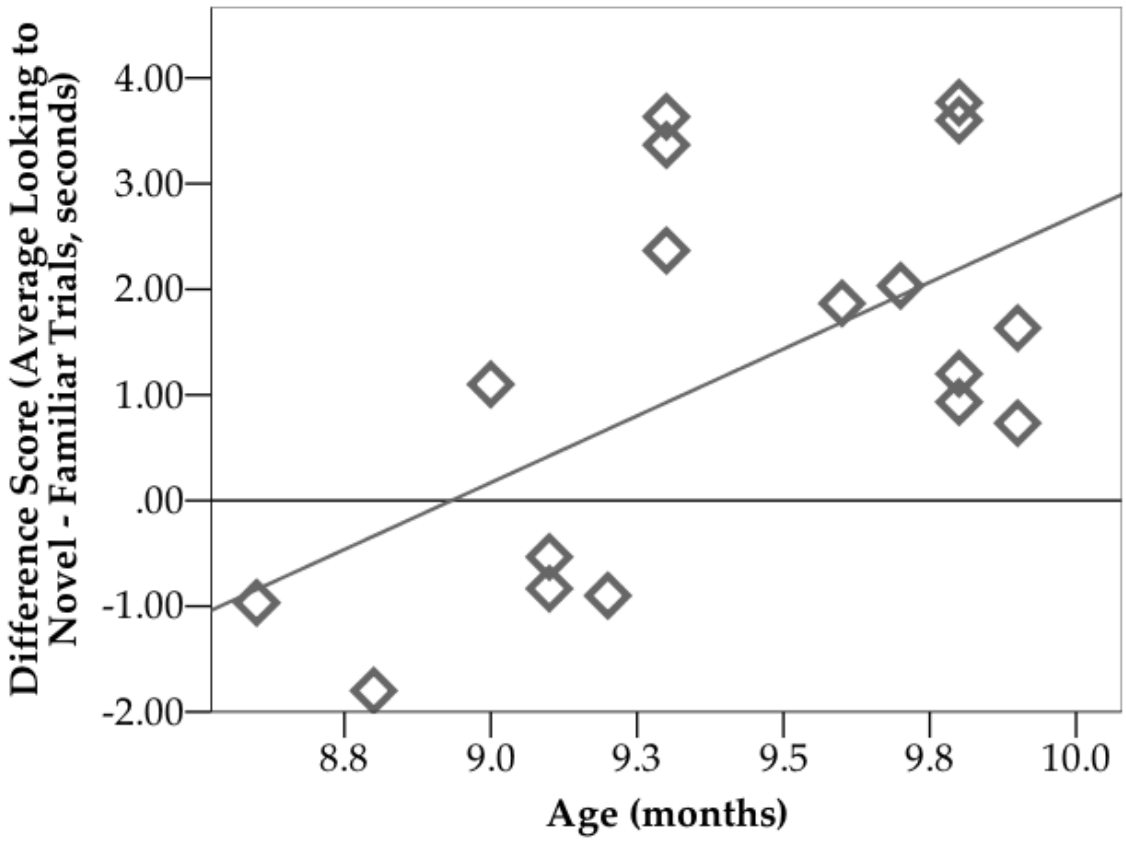


Figure 4.10: Habituated infants in the auditory condition have a significant correlation between Difference Score and their age with older infants exhibited a significant Novelty preference and younger infants exhibiting no significant learning.

tween these two factors suggests that the effect of age is not attributable to an effect of age on viewing time.

To further investigate the relationship between age and learning, we performed a median split of the group of Habituated infants in the Auditory modality. Infants in this group have a median age of 9.3 months. Thus, *younger infants* are 9.3 months and younger, and *older infants*, older than 9.3 months. We then tested whether these groups show evidence of learning (significant difference in looking in seconds to novel and familiar test trials). Younger infants show no

evidence of learning in parametric ($t(8) = 0.88$, $p = 0.41$, $d = 0.39$) and nonparametric tests (5 of the 9 infants exhibited a novely bias, $Z = -0.77$, $p = 0.44$). Older infants did show evidence of learning in both types of tests (all 8 infants have a novelty bias, Wilcoxon signed-rank test, $Z = -2.5$, $p = 0.012$; $t(7) = 4.9$, $p = 0.002$, $d = 1.37$; Figure 4.11). Thus, we found that auditory SL younger infants in our sample do not learn under current conditions while older infants do.

To compare these results to younger infants in the Visual condition, we created a group of younger infants using the same measures as for the Auditory modality. We found that in the Visual condition, infants aged 9.3 months and younger are able to learn as measured by both parametric and non-parametric tests ($t(11) = -3.89$, $p = 0.003$, $d = -0.98$; 11 of 12 infants showed a familiarity preference, $Z = -2.82$, $p = 0.005$; Figure 4.11). We found no significant difference in viewing time across modalities at this age, in seconds ($t(19) = -1.11$, $p = 0.28$, $d = -0.51$) or in number of tokens perceived ($t(19) = 1.10$, $p = 0.29$, $d = 0.48$. Thus, we found evidence that younger infants, from the current sample, are able to learn in the visual but not the auditory modalities under the same learning conditions. Additional analyses revealed no differences in Difference scores by Experimental location, Gender, Bigram order or Test Trial order in either modality condition.

4.4 Discussion

We compared statistical learning of auditory and visual stimuli in infants aged 8- to 10-months. To our knowledge, this is the first comparison of statistical learning across perceptual modalities in infancy. There are three main findings:

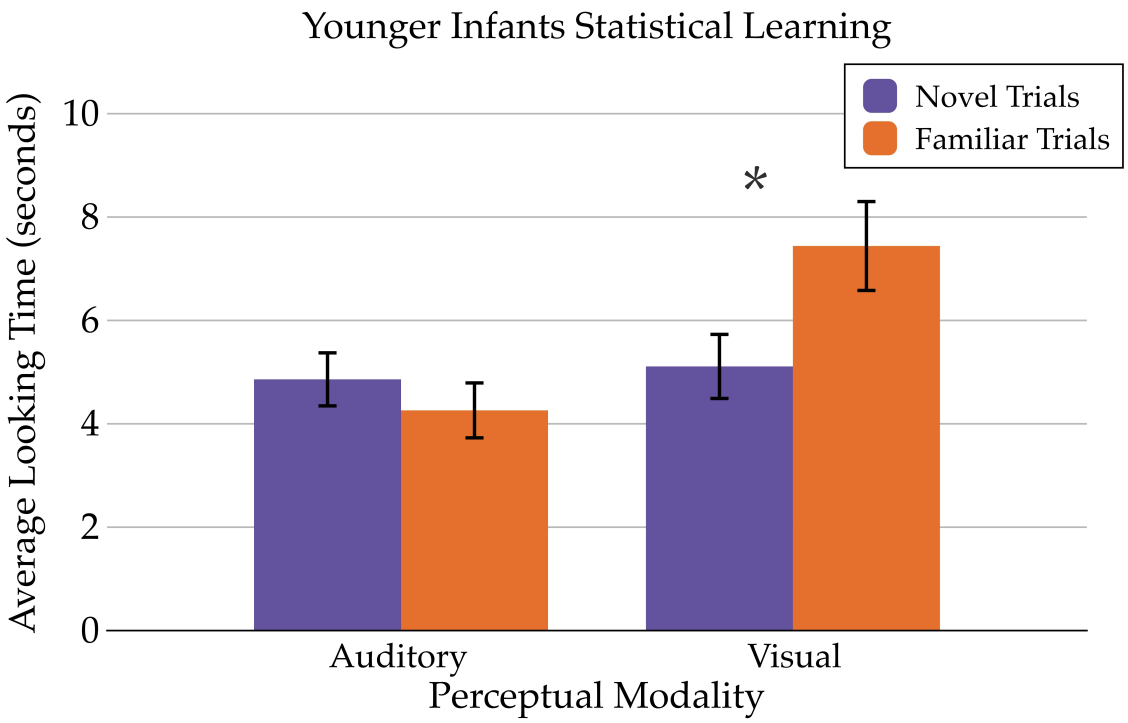


Figure 4.11: Examination of statistical learning of infants 9.3 months and younger across perceptual modalities.

1) overall, the magnitude of statistical learning is equivalent across perceptual modalities; 2) however, we find evidence for an earlier availability of statistical learning in the visual modality; and 3) although we find evidence for learning in both modalities, infants exhibit different directions of preference at test depending on modality of exposure. Taken together, these findings suggest that while the mechanisms supporting statistical learning are domain-general, they are not amodal and abstract, instead, they are sensitive to perceptual identification of statistical information. In addition, there is evidence that statistical learning is not developmentally invariant. Specifically, we find auditory SL changes with age and some evidence for earlier availability in the visual modality for the current paradigm. However, we find no evidence for changes in visual SL for the age range investigated. Thus, these results suggest differences in developmental trajectories of SL across perceptual modalities. Here, we discuss these findings

in more detail.

4.4.1 Evidence for Statistical Learning with Faces

To compare SL across auditory and visual modalities, we employed perceptually complex, biologically-relevant, and familiar stimuli: faces and speech, respectively. While previous studies have found auditory SL using speech stimuli (e.g., Saffran et al., 1996), the current results are the first evidence of SL in a stream of faces. This is an impressive ability given the complexity as well as the perceptual similarity of these stimuli: all faces were novel, smiling, Caucasian female faces (Figure 4.1). Moreover, after the first half of the first postnatal year, infants have developed specialized processing for faces (vs. other objects: De Haan & Nelson, 1999; vs. other race faces: Kelly et al., 2007). Thus, we find that despite having specialized perceptual processing of faces, infants are still sensitive to the statistical information across face presentations. [2]

4.4.2 Comparing Auditory and Visual SL

Current results indicate both similarities and differences in statistical learning from auditory and visual stimuli. In general, we find evidence that the magni-

[2]The current study employed all Caucasian faces, and while the majority of infants were also Caucasian, infants were recruited from multiple races. Despite evidence for perceptual narrowing for "other race" faces for infants in our age range, we have no evidence that this had an impact on our results. Exposure to non-native faces preserves discrimination abilities to 9 months (Pascalis et al., 2005). Infants were recruited from Ithaca and surrounding areas and from the outlying areas of Syracuse. The 2006 Census Bureau (2007a, 2007b) reported that all of these areas have a high proportion of Caucasian residents. Thus, infants will have had sufficient perceptual experience with Caucasian faces suggesting that perceptual narrowing would not yet have occurred.

tude of learning does not differ across perceptual modalities. This result holds even after controlling for generally longer looking at visual stimuli at test and for differences in the rate of stimulus presentation required to elicit learning in auditory vs. visual modalities. Thus, overall, it does not appear as if infants show a stronger ability to differentiate novel and familiar test trials in one perceptual modality over another. This is a notably different result from studies with adult learners where we find that auditory SL tends to be greater than visual SL given similar experimental conditions (Conway & Christiansen, 2005; Emberson et al., 2011; Saffran, 2002). Thus, we find evidence for a different pattern of SL across perceptual modalities for infants and adults. This suggests changes in the mechanisms of SL with development and a possible interrelationship between perceptual experience and SL.

4.4.3 Superiority of visual SL in younger infants

We found a significant correlation between age and learning in auditory SL. Specifically, we find that infants younger than 9.3 months show no evidence of learning in the auditory modality but show robust learning in the visual modality. There was no difference in viewing time across modality conditions, in either seconds or tokens, for younger infants. These results suggests greater efficacy of visual SL prior to 9.3 months.

Thus, we find that the youngest infants, tested in the current study, have the opposite pattern of learning ability across visual and auditory modalities than what has been established in adults. Why we find evidence that visual SL is superior to auditory SL at this stage of development is an important question

for future investigation. One possibility has to do with differences in visual input across development. The differences in auditory and visual processing have been conceptualized using an auditory:visual::temporal:spatial analogy where auditory information is preferentially processed in relation to temporal important and visual information is preferentially processed in relation to spatial information. There is evidence that this analogy applies to auditory and visual SL. Saffran (2002) and Conway and Christiansen (2009) found that including spatial information in visual input increased visual SL. This effect is intuitive as visual input is often spatially arrayed and suggests some correspondence between the nature of perceptual input and statistical learning abilities. Many studies have found auditory superiority in learning when SL streams are presented sequentially in both modalities (see Emberson et al., 2011, for an in-depth discussion on the relationship between spatial and temporal processing in auditory and visual SL). However, recent work has shown that the pattern of visual input is markedly different for infants. Smith et al. (2011) placed head-mounted cameras on infants aged 17 to 19 months and their caregivers while they played with a number of objects (see Figure 4.12). Relevant to the current work, they found that the infants field of view was often dominated by single objects in succession while the adults field of view often contained visual information from a number of objects. Thus, compared to adults, the infants everyday visual input may be similar to the visual sequences presented in the current study: sequential presentations of single objects. One possibility is that visual SL abilities in the current type of paradigm shift as visual input shifts from infant to adult views.

The visual SL superiority effect in younger infants is supported by the lack of learning in the auditory modality (see Figure 4.11). Previous studies have found statistical learning in the auditory modality at ages younger than 9 months (e.g.,

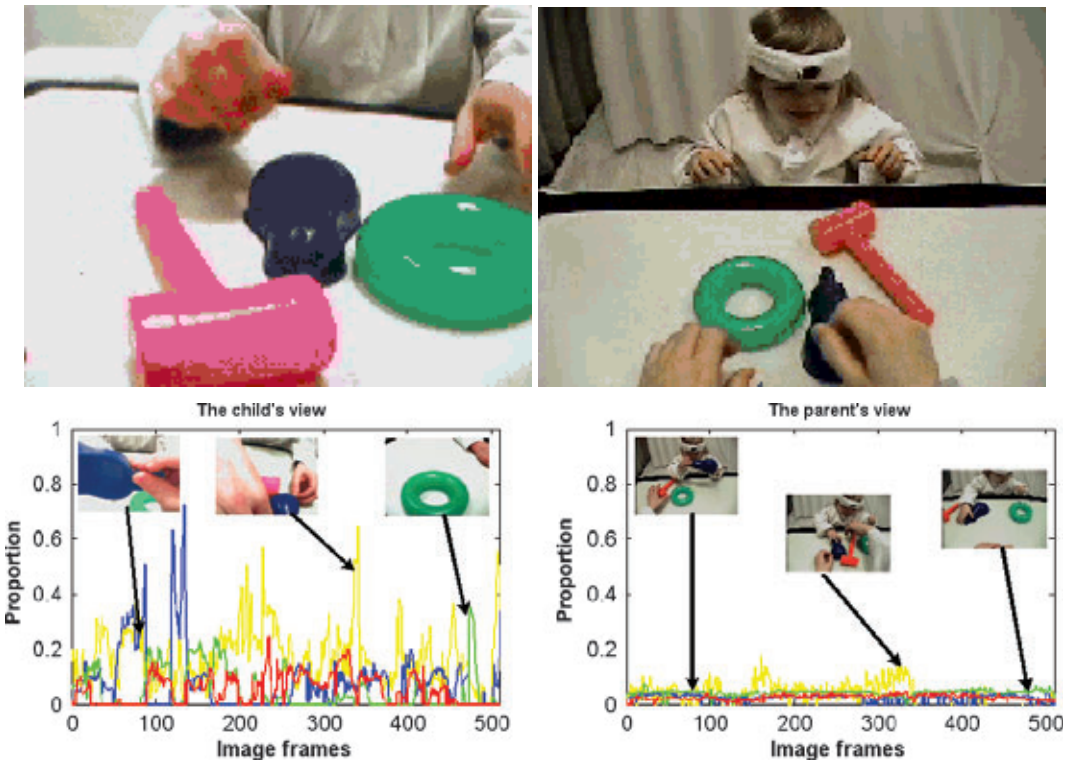


Figure 4.12: Top Panel: A child's (left) and their caregiver's (right) point of view for the same event. Pictures from Chen Yu, Indiana University. Bottom Panel: A time series of the visual input for a child-caregiver dyad (left and right, respectively). The proportion of the field of view occupied by individual objects (green, red, blue) and body parts (yellow) is shown for each time point. From Smith et al. (2011).

.

Saffran et al., 1996; Aslin et al., 1998; Thiessen et al., 2005). Infants who habituated in the Auditory condition heard, on average, 150 syllables (SD = 76 tokens). This is equivalent to 75 bigrams, or 25 presentations of each bigram. Comparing this amount of exposure to previous studies in this age range, we find that studies eliciting evidence for auditory SL employed a greater repetition of stimuli (Saffran et al., 1996; Pelucchi et al., 2009, Table 4.1). These studies also presented auditory stimuli at a faster rate of presentation. Either of these factors could explain why we fail to find auditory SL in younger infants. Notably, Thiessen et al. (2005) employed a similar rate of presentation and amount of exposure

and they found no evidence of learning. Thus, given a comparison with previous studies, it is not surprising that auditory SL was not successful for younger infants.

It is notable that older infants can learn in the auditory modality given the same experimental conditions under which younger infants cannot. These results strongly suggest that auditory SL is not developmentally invariant. A developmental shift in auditory SL of speech stimuli around 9 months is significant given the important changes in language processing occurring at this age (e.g., Bates, Bretherton, & Snyder, 1998). Thus, SL abilities as pertaining to speech stimuli may be inter-related to current language processing abilities: with SL supporting new language abilities and previous language knowledge further supporting increased SL abilities. Knowledge that speech tends to be structured by relevant and reliable statistical information might support greater engagement of learning mechanisms (for examples for this concept in the word learning literature see Smith, 2000). It is important for future research to examine whether a similar developmental shift is found in non-speech stimuli and whether this shift is related to individual developmental trajectories of language abilities.

4.4.4 Opposite Direction of Preference for Auditory and Visual SL

Another notable difference between visual and auditory SL is the direction of preference for test stimuli: infants receiving visual tokens exhibited a signif-

[3]Exposure was calculated per unit of structure (i.e. each word as in Saffran et al., 1996; each bigram in the current study; each bigram or trigram in Thiessen et al., 2005)

Table 4.1: Comparison of Rate and Exposure across Auditory SL studies

Study	Age (months)	Rate (ms)	Exposure [3]	Outcome
Current	9.1	500	25	Learning greater than 9.4 months
Thiessen, Hill & Saffran, 2005	8	400	24	No learning for AD speech
Saffran, Aslin & Newport, 1996	8	222	45	Learning
Pelucchi, Hay & Saffran, 2009	8.5	167	45	Learning

icant familiarity preference while those receiving auditory tokens exhibited a significant novelty preference. While novelty preferences are more common, familiarity preferences have been reported in previous SL and related studies (auditory SL: Saffran & Thiessen, 2003; Thiessen et al., 2005; auditory word segmentation: Jusczyk & Aslin, 1995). Studies reporting a familiarity preference have argued that a systematic preference, rather than a particular direction of preference, is the essential evidence for learning (Saffran & Thiessen, 2003; Thiessen et al., 2005). Thus, we interpret both the visual familiarity preference and the auditory novelty preference as evidence of SL.

However, opposite directions of preference emerging from largely similar experimental methodologies does suggest differences in auditory and visual SL. Moreover, the few previous investigations of visual SL in infancy have consistently reported novelty preferences (Kirkham et al., 2002; Fiser & Aslin, 2002). We examine three possible explanations for this result and their implications for the underlying mechanisms of this learning ability: 1) visual SL is more difficult than auditory SL; 2) there is an inherent familiarity preference to faces for this age range; 3) distinct cognitive mechanisms support SL across modalities.

The most common interpretation of direction of preference is based on Hunter and Ames (1988) where a familiarity preference precedes a novelty preference during stimulus encoding. According to this model, the same cognitive mechanism can give rise to different directions of preference based on the difficulty of the task. There is evidence that this model might explain directions of preference in SL tasks: Thiessen et al. (2005) found a familiarity preference in an auditory SL experiment (Exp.1), but elicited a novelty effect by making the task easier through increasing both the age of their sample and the amount of

exposure they received (Exp.2).

An explanation for the current results based on the Hunter and Ames model would imply that visual SL was harder than auditory SL because we found a familiarity effect for the visual condition and a novelty effect for the auditory condition. Thus, if a single mechanisms underlies both visual and auditory SL, this mechanism would have to have decreased efficacy for visual SL compared to auditory SL to result in different directions of preference.

However, examination of the magnitude of learning suggests a different picture: we found that the magnitude of learning is equivalent across modalities even after accounting for differences in rate of presentation suggesting no differences in the efficacy of learning. Thus, the application of the Hunter and Ames model to explain the different directions of preference is difficult to reconcile with results suggesting a similar magnitude of learning across perceptual modalities. [4]

A second possibility is that specialized processing of faces at this age drives a familiarity preference. We employed auditory and visual stimuli that are both subject to substantial developmental changes in the first post-natal year facilitated by perceptual experience (speech and faces). It is possible that SL is affected by the specialized processing of these stimuli resulted in differences in learning mechanisms. The present study is the first to investigate SL with faces. Studies of similar age ranges suggest that a novelty preference is more common for face processing after the first half of the first post-natal year. However,

[4]It is important to note that the number of stimuli perceived during habituation was greater for infants in the auditory modality. The Hunter and Ames model also assumes that greater exposure results in deeper encoding and thus an increased likelihood of a novelty preference. However, this explanation is still inconsistent with the magnitude of learning results as the magnitude of learning is not different even when considering looking times in terms of the number of stimuli perceived.

Lewis (1969) found that infants at 9 months of age show a familiarity preference for typical faces as opposed to scrambled or Cyclops faces. The visual schema of a face could be viewed as a learned structure supported by the visual statistics of face components and thus infants preferred stimuli that conform to familiar statistics rather than novel face-like stimuli that violate this structure. Thus a possible explanation for the visual (face) SL familiarity preference is that infants prefer familiar *sequential statistics* of faces in the same way that infants prefer the *structural statistics* of typical faces. Future studies would be needed to further investigate whether infants have a preference for familiar statistics for face stimuli in general.

A third possibility is that visual and auditory SL are supported by distinct mechanisms. This is an important area of future study as very little is known about the mechanisms underlying SL across perceptual modalities. Since previous research has not established that visual and auditory SL are supported by the same neural mechanisms, many alternative possibilities exist. However, a number of studies suggest the possibility is that visual SL is more reliant on the hippocampus to bind across successive visual experiences while auditory SL does not depend upon this learning system. Familiarity preferences are consistently reported in studies examining relational memory for complex visual stimuli. For example, Richmond and Nelson (2009) showed 9-month-olds faces paired with scenes. During test, infants saw the same scenes paired with three faces: the one previously paired with this scene and two were familiar but not previously paired with this scene. Infants preferentially looked at the face previously paired with the scene (see also Rose, Gottfried, Melloy-Carminar, & Bridger, 1982). A similar familiarity preference is also reported in adults. Moreover, a lack of familiarity preference in MTL amnesiacs sug-

gests that this task is supported by the hippocampus (Hannula, Ryan, Tranel, & Cohen, 2007). Relatedly, neuroimaging studies in adults have established that visual SL is supported in part by activity in the hippocampus (Turk-Browne et al., 2009, 2010) consistent with the role of this learning system in binding across experiences to support integrated memory (Amso, Davidson, Johnson, Glover, & Casey, 2005; Shohamy & Wagner, 2008). While the hippocampus has been implicated in some areas of language learning (e.g. word-meaning mappings; Rodríguez-Fornells, Cunillera, Mestres-Missé, & Diego-Balaguer, 2009), it has not been implicated in auditory SL even in participants as young as 6 years of age. Instead, auditory SL appears to be supported by regions in the temporal, frontal and dorsal parietal cortices with changes over development (McNealy, Mazziotta, & Dapretto, 2006, 2010; McNealy et al., in press). Thus, support of visual and auditory SL by different neural mechanisms could explain the different directions of preference at test and differences in developmental trajectory. Moreover, if visual SL with complex stimuli is more reliant on the hippocampus, as outlined here, the familiarity preference in visual SL is more consistent with relational memory tasks which also rely on the hippocampus.

4.4.5 Exposure vs. Attention during Habituation

To our knowledge, this is the first study of auditory SL to use infant controlled-habituation. Previous studies have used fixed habituation: a methodology that allows for all infants to receive the same amount of exposure to the statistical regularities regardless of directed attention or an infants habituation status and preferences. Having used infant-controlled methodologies in both modalities, we were able to examine the inter-relation of amount of exposure to statistical

regularities or viewing time during habituation, the reaching of the habituation criterion and SL across perceptual modalities. We found that Non-Habituated infants show only marginal evidence of learning in both modalities (and again with different directions of preference). This result is notable because infants who did not habituate looked significantly longer during habituation and thus received greater exposure to the statistical information than Habituated infants. This finding suggests that statistical learning may not simply be driven by cumulative, passive exposure (i.e. greater exposure to statistical information results in more SL). Instead, directed attention and, relatedly, habituation play an important role in infants SL abilities. This result is consistent with an emerging view that overt looking, measured in amount of time looking, is not a direct measure of attention in infancy (see M. Johnson, Posner, & Rothbart, 1994, for a dissociation of eye gaze and attention in 4-month-old infants). Moreover, this finding parallels emerging evidence that adult learners benefit and perhaps require attention to stimuli endowed with statistical information in order to learn (Emberson et al., 2011; Toro et al., 2005; Turk-Browne et al., 2005; however, see Saffran et al., 1997).

We also found a correlation between viewing time during habituation and difference scores–infants who looked longer tended to have a stronger novelty preference–only for Habituated infants and only for those receiving auditory experience. This result also suggests that habituation, rather than amount of exposure, supports SL in infants. However, it is surprising to find no parallel relationship in the visual modality. Future research is required to more clearly investigate the relationship between directed attention and learning across perceptual modalities.

Modality-Specific Effects of Rate of Presentation on SL. We found that when auditory and visual streams are presented at the same rate (1 stimulus/second), there is evidence of visual learning only. However, when auditory stimuli are presented at double the rate (2 stimuli/second), we find clear evidence for auditory learning. The same interrelationship between rate and auditory SL has been reported in adult learners (Emberson et al., 2011). Moreover, Arciuli and Simpson (2011) a constant effect of rate in visual SL with age in children. In other words, they find that the effect of rate does not interact with effects of age. While more systematic studies are required, these findings suggest that the interaction of perceptual modality with presentation rate and subsequently SL are conserved throughout development. If learning across perceptual modalities were indeed constrained to different rates of informational input, this would provide a modality-specific bias on the patterns that can be learned given a given type of perceptual input (e.g. faces vs. speech).

In sum, previous research has suggested that SL is mediated by an abstract, amodal, developmentally-invariant mechanism. Such a view would predict that statistical learning would be equivalent both across perceptual modalities and over developmental time. We report the first evidence that auditory and visual SL outcomes are not identical even in the same learning paradigm: there are differences in the direction of preference across modalities and the relationship between viewing time and learning outcomes. Moreover, results suggest the presence of different developmental trajectories in SL across modalities: first, visual SL is more robust than auditory SL before 9.3 months; second, only auditory SL shows a positive correlation with age. Finally, the relationship between learning outcomes across vision and audition consistently differs from the pattern observed in adults. Future research needs to establish what experience or

maturational changes support shifts in the pattern of learning, however, there is some suggestion that changes in learning could mirror concurrent changes in cognitive development (i.e. changes in auditory SL at 9 months) and perceptual input (i.e. differences in visual input from infancy to adulthood). Overall, the current study presents a novel view on the mechanisms that support the ability to learn from structure in the environment: incidental statistical learning abilities in infancy are intertwined with perception, attention, and other cognitive abilities.

CHAPTER 5

INVARIANCE FROM VARIABILITY: THE MEDIAL TEMPORAL LOBE SUPPORTS CHANGES IN OBJECT PERCEPTION

The content of this chapter is currently under review and was written with Dr. Dima Amso at Brown University. See Emberson and Amso (in preparation).

We used a combined fMRI/eye-tracking approach to examine the mechanisms involved in learning from visual experience to segment a novel, occluded object in a scene. Previous research has suggested a role for effective visual sampling and prior experience in the development of mature object perception. However, it remains unknown how the naïve system integrates across variable experiences to induce perceptual change. We generated a Target Scene in which a novel occluded Target Object could be perceived as either disconnected or completed. We presented one group of participants with this image in concert with variable but regular visual experience: three Paired Scenes consisting of the same Target Object in variable locations, orientations, and states of occlusion. A control group was presented with similar Paired Scenes that did not incorporate the Target Object. We found that, relative to the control group, participants in the Training condition were significantly more likely to change their percept from "disconnected" to "completed", as indexed by pre- and post-test performance. In addition, gaze patterns during Target Scene inspection differed as a function of regular, variable object exposure. Our neuroimaging findings provide compelling evidence that the hippocampus, together with the ventral visual pathway, is involved in binding across variable visual experiences to ultimately engage effective sampling and perceptual change. We propose a role of these mechanisms in catalyzing perceptual development

5.1 Introduction

The mechanisms of object perception and recognition have received considerable scientific attention. Discovery in this domain holds substantial promise for informing one of the most important questions in cognitive and brain sciences: How do we construct and act on an enduring representation of the external environment? There are a variety of research avenues and levels of analysis appropriate for addressing this question; they span everything from the size of receptive fields along visual pathways to philosophical discussions about the origins of object concepts. Here we focus on mechanisms of information acquisition. Research into the development of object perception suggests a role for both effective sampling and prior experience in mature object perception. This work exposes the mechanistic nature of the interaction between sampling and integrating across variable experiences during learning to segment a novel object.

A hallmark of visual object processing is perceptual completion: the perception of an occluded object as a complete whole despite visual evidence to the contrary. A series of studies have determined a role for developing sampling, via selective attention, mechanisms in the emergence of this skill (Amso & Johnson, 2006; Schlesinger, Amso, & Johnson, 2007). Sampling may serve to support the extraction of object feature correlations in the service of efficient perception and recognition (Bhatt & Quinn, 2011). Previous research has found that 3-month-olds who indicated perceptual completion of an occluded object also targeted scans and fixations to the object parts in an otherwise cluttered scene (S. Johnson, Slemmer, & Amso, 2004). Similarly, infants gaze patterns to faces are related to their ability to discriminate facial emotions (Amso, Fitzgerald,

Davidow, Gilhooly, & Tottenham, 2010). While it is possible that an active percept drives sampling in these studies, it has been demonstrated that infants who provide evidence of perceiving occluded object unity also showed evidence for visual search behavior indicative of attention-guided eye movements (Amso & Johnson, 2006). This was taken as evidence that sampling may precede unity perception. Moreover, neuronal firing in the frontal eye fields, central to visual sampling, has been shown to precede firing in regions that contribute to object recognition, namely inferotemporal cortex (Monosov, Sheinberg, & Thompson, 2010). In spite of this evidence, the findings from the naïve infant system remain correlational and a question key to the development of object perception and recognition remains open. How does the visual system learn what to sample in variable and cluttered scenes?

Clues are offered from literatures that suggest that visual experience also plays an essential role in the development of object perception (E. J. Gibson, 1969; Kersten, Mamassian, & Yuille, 2004). Infants as young as 4.5 months can use variable visual experience with an object to segment a novel visual scene (Needham, Dueker, & Lockhead, 2005; Needham & Modi, 1999). Similarly, the adult visual system can develop sophisticated visual processing of a novel class of complex objects with a short training period both behaviorally (Gauthier & Tarr, 1997) and neurally (Gauthier et al., 1999). Hence, visual experience arguably supports successive object perception starting early in development and continuing across the lifespan.

The efficacy of learning from experience must in part rely on a mechanism that integrates across multiple, variable experiences of an object or class of objects. Hypothetically, such an integration process would require memory of pre-

vious episodes related to the current experience, as well as a process of binding these related instances to form a lasting knowledge structure. Previous research has found that both the basal ganglia and the hippocampus are involved in learning across successive visual experiences (Amso et al., 2005; Turk-Browne et al., 2009). While there is extensive work suggesting that the hippocampus is involved in creating *discrete* memories (e.g., Atallah et al., 2004; McNaughton & Nadel, 1989), there is also evidence suggesting that the hippocampus is involved in *integrating* across memories or experiences with overlapping input at the point of encoding (e.g. Shohamy & Wagner, 2008). Such integration and binding could support learning from successive, variable experience to change object perception.

An important consideration is how similar two learning episodes must be to be bound by a learning mechanism. Invariance is an important property of the visual system whereby objects are recognized as the same, independent of changes in perceptual information such as scale, location, and orientation (Deco & Rolls, 2004; Desimone, 1991; Rolls, 1992, 2000). This is a well-studied phenomenon that involves the entire ventral visual pathway (V1, V2, V4, and inferotemporal cortex or IT). Along this pathway, IT neuronal firing has been found to most highly correlate with conscious perception (Blake & Logothetis, 2002; Leopold & Logothetis, 1999) and to show various degrees of invariance to image transformations (Hasselmo, Rolls, Baylis, & Nalwa, 1989; Ito, Tamura, Fujita, & Tanaka, 1995; Kobatake & Tanaka, 1994; Op De Beeck & Vogels, 2000; Tovee, Rolls, & Azzopardi, 1994; Booth & Rolls, 1998).

Inferotemporal cortex is highly connected to medial temporal lobe (MTL) regions, including the hippocampus and parahippocampal cortex. Indeed, it has

recently been proposed that the MTL is involved in visual perception (Baxter, 2009; Ison & Quiroga, 2008; also see Suzuki, 2009). For example, research has suggested that the MTL has a role in the encoding of object identity (Quiroga, Reddy, Kreiman, Koch, & Fried, 2005). Moreover, single cell recordings from epilepsy patients have shown that there are category-specific neurons in the MTL (Fried, MacDonald, & Wilson, 1997; Kreiman, Koch, & Fried, 2000). That is, they respond to visual stimuli from categories of animals or places, for example. Such invariant representations may be linked to the process of forming associations between variable views of the same object. Alternatively, these categories may be a by-product of the binding process. Support for the latter comes from modeling work by Deco and Rolls (2004).

Using the principle of trace learning, a version of Hebbian learning, they showed that different views of an object that occur close together in time are associated or bound. Trace learning generated invariant neuronal responses to different transforms of an object.

We used a training paradigm with combined functional neuroimaging (fMRI)/eye-tracking methods in adults to provide insight into the mechanisms of successful online sampling and integration across variable visual experiences. In an effort to mimic the naïve visual experience, we tested adults on whether variable but regular exposure to a novel visually occluded object, embedded in varying scenes, would support change from perceiving the object as two disconnected parts to perceiving it as one occluded whole (i.e. the Training condition). In the Control condition, the same Target Scene is paired with equally complex images that did not include the varying exposure to the novel object (see Figure 5.1).

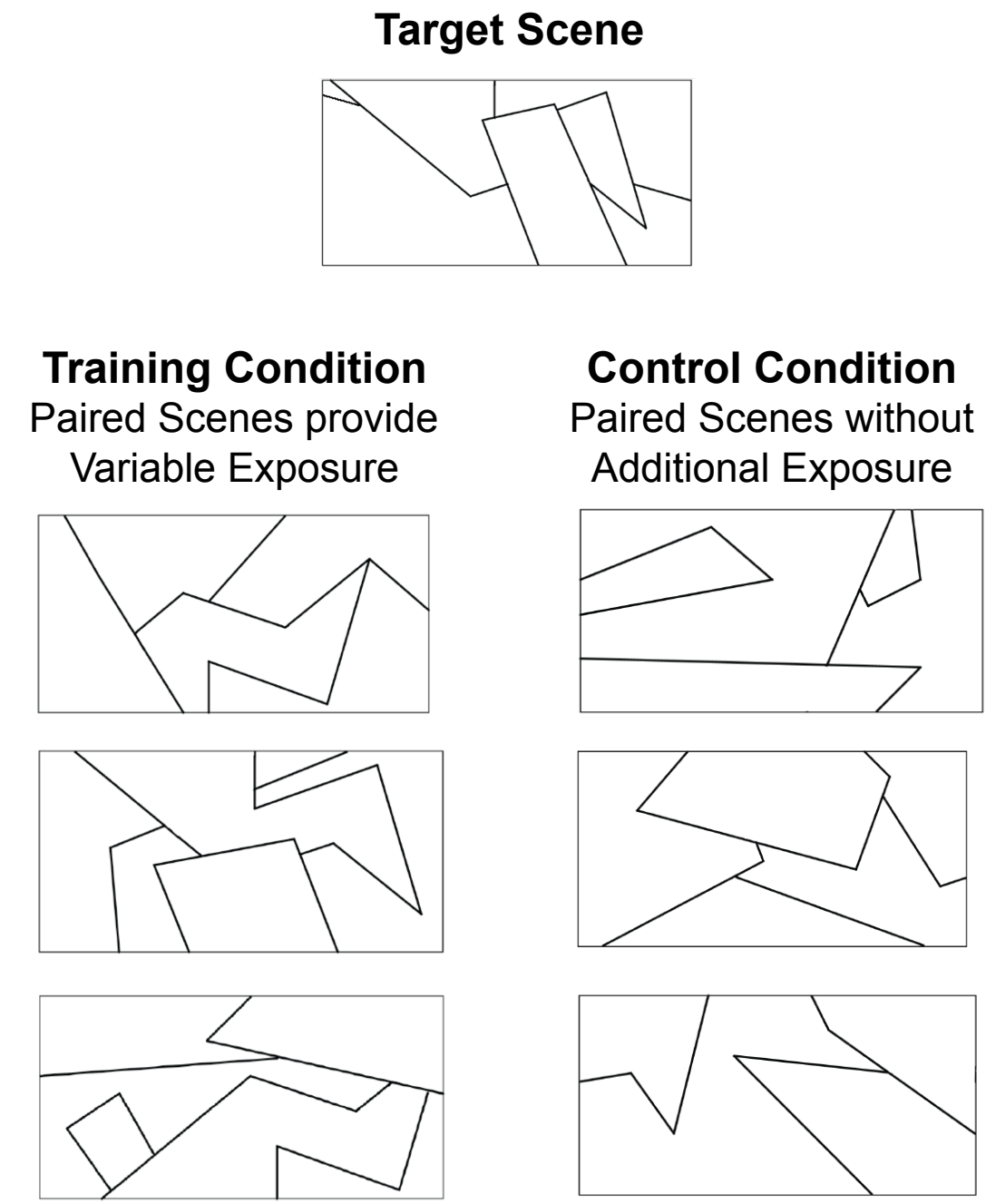


Figure 5.1: Visual Stimuli Employed in this Study. While the same Target Scene was employed across conditions, the Paired Scenes varied across condition. In the Training condition, the Paired Scenes included the Target Object in varying positions, orientations etc. (left panel), while the Control condition included Paired Scenes that were equally complex but did not include the Target Object (right panel).

This study examines the role of sampling and experiential integration in perceptual change. Based on previous research, we predict that efficient sampling, supported by fronto-parietal circuitry, will support change in object perception, perhaps through the extraction of feature correlations. Furthermore, we further predict that this sampling will be driven by effective use of regular variable exposure, as indicated by involvement of the ventral visual pathway and the MTL.

5.2 Results

5.2.1 Behavioral Efficacy of Training

A total of 61 participants composed the final sample. Behavioral participants were randomly assigned to one of two conditions (Training condition N = 20, Control condition N = 20) during passive viewing. Scanner subjects participated only in the Training condition (an additional N = 21). Both conditions provided equal presentation of the identical Target Scene. In the Training condition, however, the Target Object was repeated in each of the three additional, Paired Scenes, in various locations, orientations, and states of occlusion (see Figure 5.1).

Each participant's percept of the Target image was tested before and after exposure training. All participants were asked to indicate their perception by coloring the black and white scenes. This task allowed participants to report their exact perception without time constraints or ambiguity in their verbal responses. All participants included in the analyses reported an initial "discon-

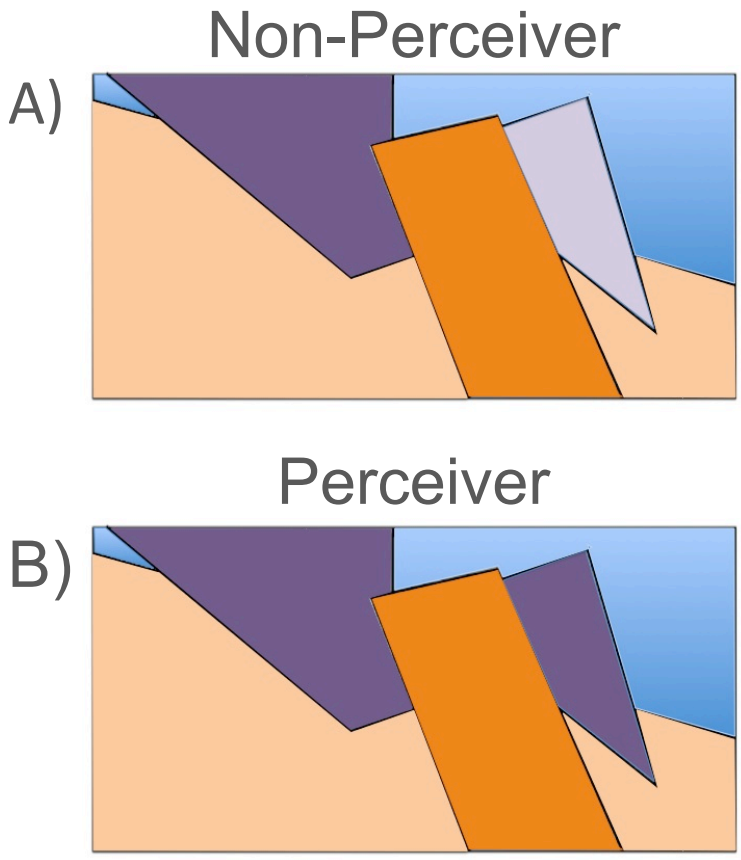


Figure 5.2: Two Sample Percepts of the Target Scene. Scenes were colored by participants such that the same object in the scene is colored the same color. The scene in panel A) depicts what would be classified as a disconnected percept, specifically, where the two surfaces that comprise the novel object are colored two separate colors. B) This scene would be classified as a connected percept because both surfaces that comprise the novel object are colored the same color depicts the object as completed behind the occluder (colored in blue in this scene).

nected" percept where the two surfaces of the novel object are colored as two separate objects and not as a single, occluded whole. We used two separate indices of post-exposure performance, a verbal and coloring task, to reliably group participants either into Perceiver or Non-Perceiver groups. Perceivers identified the Target Object during both tasks as complete–a change from their initial "disconnected" percept–while Non-Perceivers persisted in their initial disconnected percept (Figure 5.2).

Training with variable regular exposure to the Target Object is an effective means of driving perceptual change; that is participants in the Training condition were more likely to be Perceivers than those in the Control condition. In the Training condition, 65% were Perceivers while 35% were Non-Perceivers. In contrast, the Control condition yielded only 20% Perceivers while 80% remained Non-Perceivers (Figure 5.3). A Pearson chi-squared test with two factors (Condition: Training vs. Control) and (Group: Perceiver vs. Non-Perceiver) established a non-uniform distribution across Conditions: $\chi^2(1, N = 40) = 8.29$, $p < 0.005$. Thus, we can conclude that our training was successful. Repeated viewing of the Target Object, in variable orientations, locations, and states of occlusion, resulted in successful change in perception of the occluded Target Object from broken to complete. Moreover, it suggests that repeating the same Target Scene in isolation, as in the Control condition, is not sufficient to power such a perceptual shift. Repeating this analysis with neuroimaging participants included does not change the rejection of the null hypothesis, $\chi^2(1, N = 59) = 7.11$, $p < 0.01$. An additional chi-squared test, including factors of Group and Location, revealed no differences in the distribution of Perceivers vs. Non-Perceivers inside vs. outside the scanner, χ^2 $(1, N = 39)$ 1.232, $p = 0.267$.

5.2.2 Eye Movement Patterns

Regularity in the Training condition biases looking to the object parts. We parsed the Target Scene into four areas of interest (AOIs) and considered patterns of looking to those AOIs (Figure 5.4) as a function of exposure condition. The Control condition was only tested behaviorally in the lab. In order to meet the assumptions of equal *Ns* per condition, we included only data collected in the

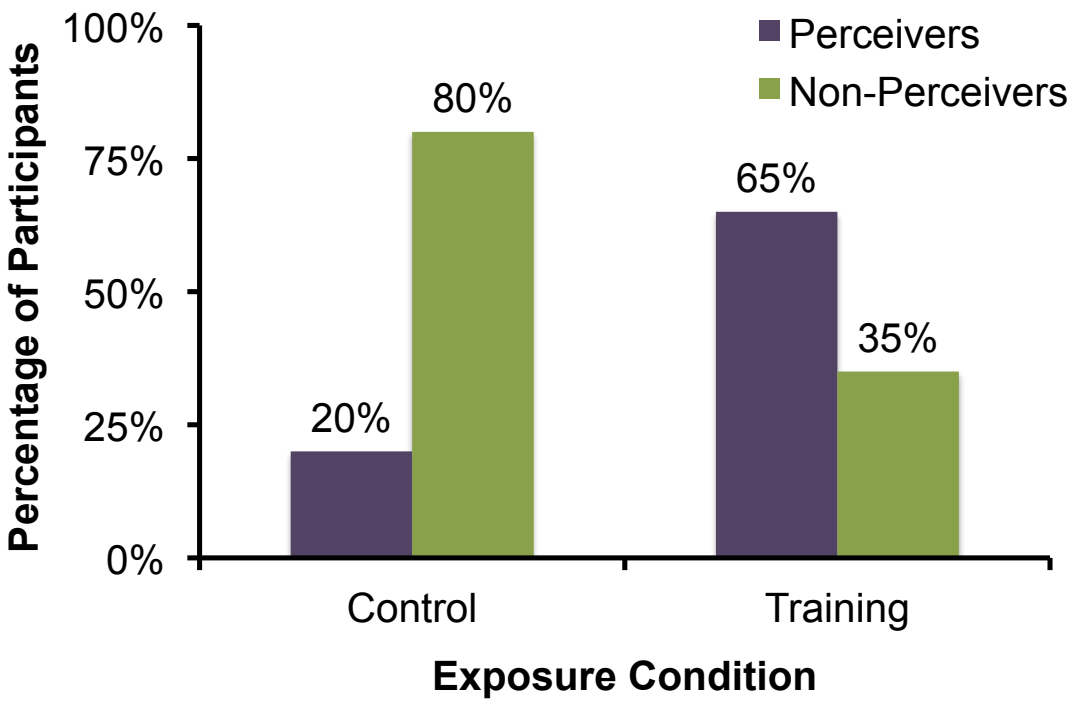


Figure 5.3: Percentage of Perceivers to Non-Perceivers in the Training and Control Conditions. Perceivers identified the Target Object during both tasks as complete, while Non-Perceivers identified it as disconnected. Chi squared statistics support a difference in the distribution of Perceivers and Non-Perceivers across these conditions.

lab for this comparison. Table 5.1 presents the proportion of looking for each AOI for all participants. We conducted a proportion of fixation duration (see Eye Tracking Preprocessing section below) per Target Scene AOI 4 (Object-1, Object-2, Occluder, and Background, Figure 5.4) x Exposure Interval 6 (data binned across task to index change over exposure) x Condition 2 (Training x Control) x Post-Test Group 2 (Perceiver x Non-Perceiver) analysis of variance on the behavioral data (N = 40). We found a main effect of Condition, $F(1, 36)$ = 9.78, $p < 0.005$. However, this subtle effect (difference in proportion of looking duration = 0.004) did not interact with Group. The analysis also yielded a main effect of AOI, $F(1.3, 108)$ = 98.7, $p < 0.001$, and an AOI x Condition interaction $F(1.3, 108)$ = 16.9, $p < 0.001$. Follow-up tests show that there is a greater pro-

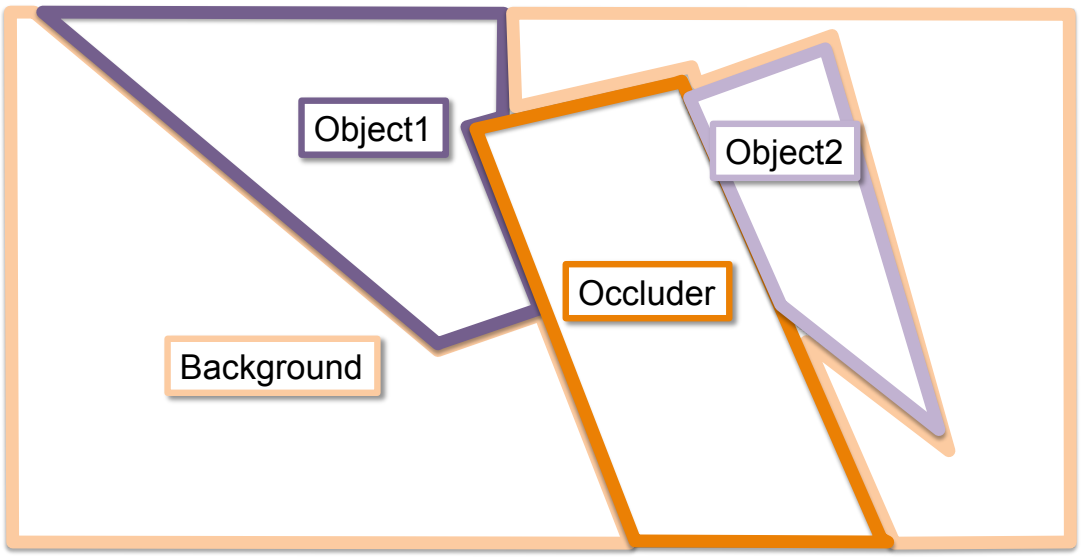


Figure 5.4: Areas of Interest (AOIs) for the Target Scene. The Target Scene was segmented into four AOIs: two separate AOIs for the two surfaces of the target novel object (Object-1 and Object-2), the occluding object separating the two object AOIs (Occluder) and finally, all surfaces of the background were treated as a single AOI (Background). All fixations during Target Scene exposure were localized to one of these AOIs. Fixations outside these AOIs were not included in subsequent analyses.

portion of looking allocated to the Object parts in the Training condition and a greater proportion allocated to the Occluder and the Background in the Control condition (see Figure 5.5, all $ts > |3.6|$, $ps \leq 0.001$, Bonferroni-corrected alpha set to .05/4). We include Figure 5.6 to specifically illustrate the striking difference in sampling between Non-Perceivers in the Training and the Control conditions. Regular but variable exposure in the Training condition powers efficient sampling even when the subject does not ultimately connect the object parts.

Sampling differences between Perceivers and Non-Perceivers in the Training condition are modest. An ANOVA of the Training condition (N = 39; Figure 5.7) across AOI, Training Interval, Post-test Group (N = 22 Perceivers, 17 Non-Perceivers) and Location (lab vs. scanner) revealed a main effect of AOI, $F(1.95, 105) = 54.5$,

Table 5.1: Proportion of looking to areas of interest in the Target Scene.

Area of Interest	Group	Proportion of Looking	*SD*
Object-1	Structured: Perceivers	0.174	0.077
	Structured: Non-Perceivers	0.262	0.156
	Unstructured: All Participants	0.025	0.019
Object-2	Structured: Perceivers	0.130	0.071
	Structured: Non-Perceivers	0.089	0.065
	Unstructured: All Participants	0.018	0.014
Occluder	Structured: Perceivers	0.515	0.132
	Structured: Non-Perceivers	0.439	0.153
	Unstructured: All Participants	0.665	0.169
Background	Structured: Perceivers	0.181	0.088
	Structured: Non-Perceivers	0.211	0.118
	Unstructured: All Participants	0.290	0.147

$p < 0.001$. Participants in the Training condition looked most at the Occluder and they look least at the Object-2 surface (all tests with Occluder and Object-2, $ts(38) > |3.6|$, $ps \leq 0.001$, Bonferroni-corrected alpha $p = .008$). This was qualified by a marginally significant AOI by Group interaction, $F(1.92, 105) = 2.70$, $p = 0.08$. We used planned comparisons to examine informative differences in looking distributions within each Group and found them to be remarkably similar across Groups (Figure 5.7). One notable difference is that Perceivers show a more even distribution of looking between the Object surfaces, while Non-Perceivers show a significant difference in looking between the two Object surfaces, $t(16) = 3.9$, $p = 0.001$. Specifically, Non-Perceivers look more at Object-1, $t(37) = -2.25$, $p = 0.031$ than at Object-2, which contributed to the main effect of AOI described above. A more even distribution, exhibited by Perceivers, might be related to extracting feature correlations across object parts and Training images. We also find that Non-Perceivers look more at Object-1 than Perceivers, $t(37) = -2.25$, $p = 0.031$ and Perceivers look marginally more at Object-2, $t(37) = 1.86$, $p = 0.071$.

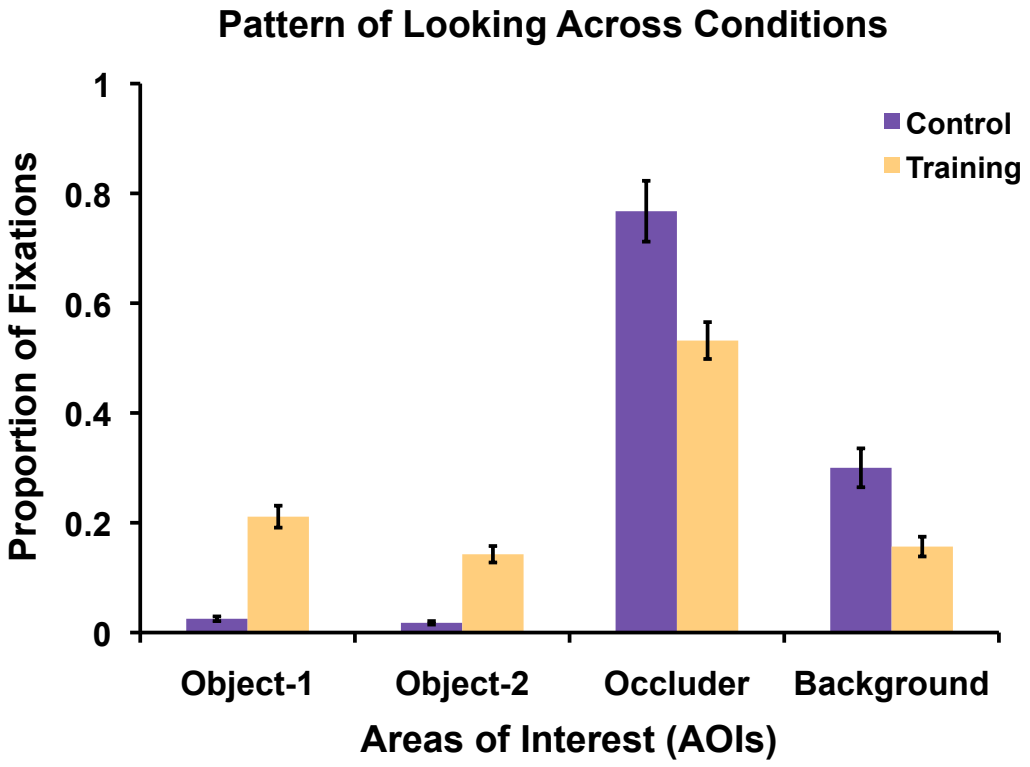


Figure 5.5: Proportion of Fixations to the Target Scene in the Control and Training Conditions. Looking is presented separated in each of the four AOIs used in the eye tracking analyses. Participants in the Training condition look more to the two Object parts and less to the Occluder and Background demonstrating that variable exposure to a novel object results in different distributions of fixations to the Target Scene.

We found no main effects associated with Training Interval or Testing Location in the Training condition. There was a significant Training Interval by Group by Location interaction, $F(2.6, 175) = 3.5$, $p = 0.02$. Follow-up analyses reveal a significant Training Interval by Group interaction in the scanner, $F(2.0, 85) = 4.34$, $p = 0.02$, but not outside of the scanner, $p > 0.7$. This subtle effect is likely the result of the differences in timing across Locations. However, these differences do not seem to meaningfully or systematically reflect in the AOI distributions.

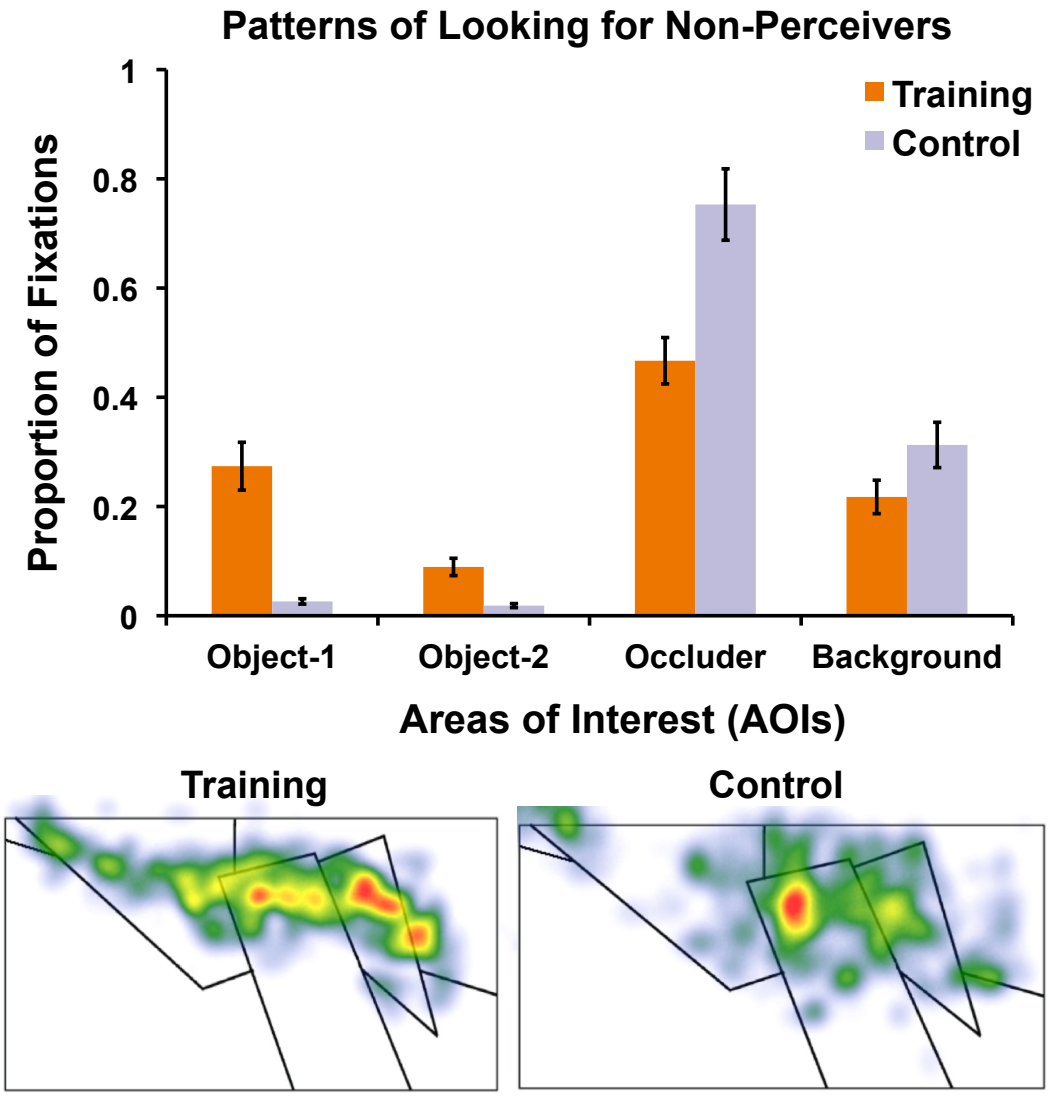


Figure 5.6: Pattern of Looking for Non-Perceivers across Exposure Conditions. Top Panel: Proportion of fixations to the Target Scene for Non-Perceivers in the Control and Training conditions. Bottom Panel: Differences in fixation patterns for two representative Non-Perceivers from the Training and Control conditions. Hotter colors are longer average duration of fixations.

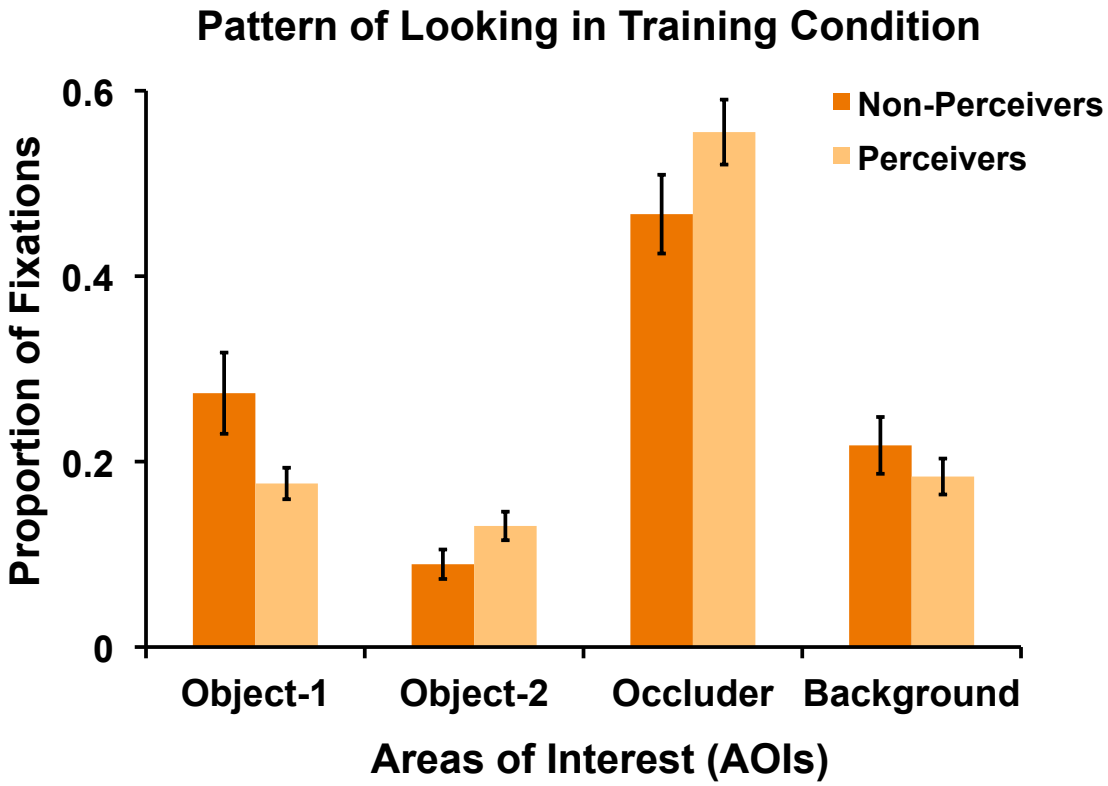


Figure 5.7: Proportion of Fixations to the Target Scene in the Training Condition by Group. Looking is presented separated in each of the four AOIs used in the eye tracking analyses. We find significant differences across Groups for Object-1 AOI only.

Looking duration proportions that are corrected for AOI size confirm these patterns. AOIs in the current task vary in their surface area and shape. If fixations were randomly distributed, the proportion of fixations to each AOI would be equivalent to their proportionate size. We examined looking patterns while controlling for differences in surface area. To this end, we calculated the proportion of fixations expected per AOI–the proportion pixels for each AOI relative to total scene size–subtracted from the proportion of fixations observed and arcsine transformed. We compared this difference score value to zero (the value at which subjects looked at AOIs exactly as would be predicted by area alone). For the Control condition, we find that participants look significantly more at

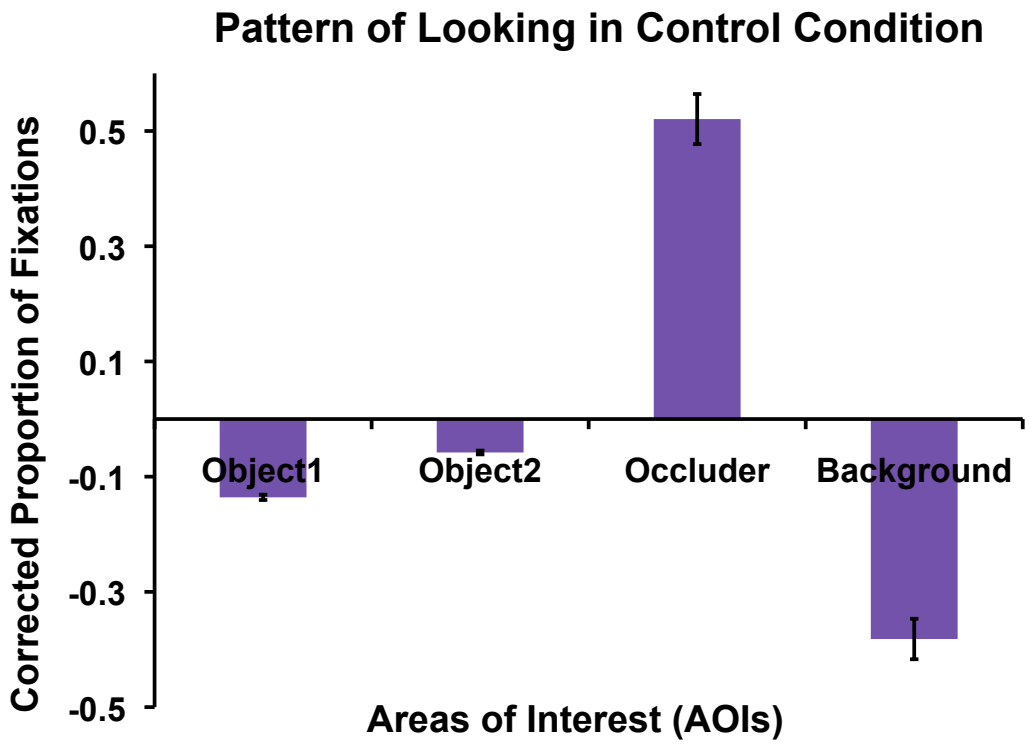


Figure 5.8: Corrected Proportion of Fixations to the Target Scene in the Control Condition. In general, participants look most at the occluder and least at the Object-1 surface. There are no differences in looking between Perceivers and Non-Perceivers in the Control condition.

the Occluder than would be expected by area size and significantly less to all other surfaces, $ts(19) = |7.8|$, $ps < 0.01$, see Figure 5.8. This is perhaps because it is centrally located in the Target Scene. We found a markedly different pattern for the Training condition (see Figure 5.9). Participants looked at the Occluder and both Object surfaces more than would be expected by chance, and look less at the Background, $ts(38) = |2.6|$, $ps < 0.01$. Perceivers in the Training condition show significantly more looking for the Object-2 surface, $t(21) = 3.7$, $p = 0.001$, than would be expected as a function of area size, while Non-Perceivers show greater than expected looking to the Object-1 surface, $t(16) = 2.6$, $p = 0.02$. These results suggest that Groups allocate their attention differently across Object sur-

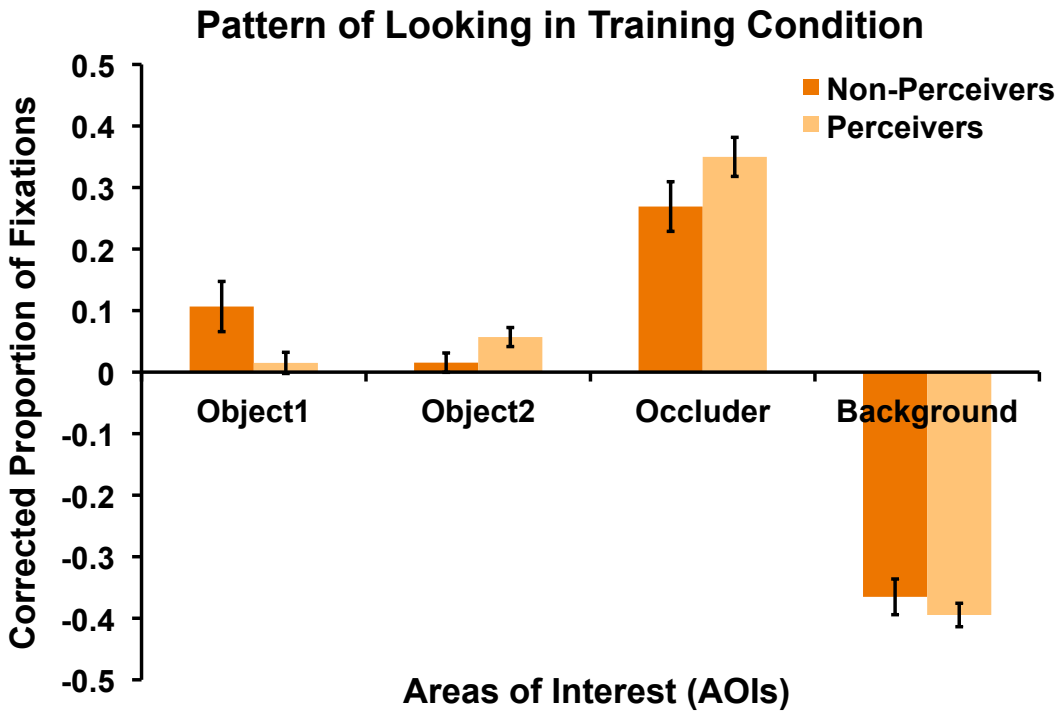


Figure 5.9: Corrected Proportion of Fixations to the Target Scene in the Training Condition by Group. Non-Perceivers look more at the Object-1 surface while Perceivers look more at the Object-2 surface.

faces.

Finally, we submitted these corrected data to the ANOVAs reported for the raw data. This resulted in virtually identical pattern of effects, confirming that these analyses are not confounded by surface area differences. The exception is an additional main effect of Training Interval for analysis comparing looking for behavioral subjects across Conditions and Groups, $F(5, 275) = 2.51$, $p = 0.03$. This effect does not interact with Group or Condition variables and does not appear to meaningfully influence with the broader pattern of results.

Taken together, variable exposure is relevant for sampling. However, differences between Perceivers and Non-Perceivers are modest, with the exception of a bias for a more even distribution of looking to the object surfaces in Per-

ceivers relative to Non-Perceivers. That is, there is not convincing evidence that sampling alone is sufficient to catalyze the shift in object perception between Perceivers and Non-Perceivers. Indeed, sampling differences in the Training relative to the Control condition in Non-Perceivers are obvious in Figure 5.6, even as the participants in the Training condition do not make the perceptual shift. We turn to imaging data to expose other relevant processing differences between groups.

5.2.3 Neuroimaging Results

FMRI data indicate important differences in learning systems and the ventral visual stream between Perceivers and Non-Perceivers. We collected fMRI data in the Training condition only and conducted a whole brain analysis for Group (Perceivers and Non-Perceivers) by scene type (Target Scene relative to the Paired Scenes – the three scenes with variable views of the Target Object; Figure 5.1 left panel). We hypothesized that the neural systems involved in relating information from Paired Scenes to the Target Scene will be modulated by Group. Regions relevant to this interaction of scene type and Group are listed in Table 5.2. We found differences in temporal lobe regions spanning portions of the fusiform gyrus (FFG), the inferior temporal gyrus (ITG) and the hippocampus. Follow-up analyses in the hippocampal ROI (see analysis details in Methods below; Figure 5.10A) revealed a significant difference between the Target and Paired Scenes in the Perceivers only (PSC: $t(8) = -2.46$, $p = 0.039$; betas: $t(8) = -2.3$, $p = 0.05$; Figure 5.10B left). Non-Perceivers showed no difference in hippocampal sensitivity to this contrast (Figure 5.10B right). This pattern in the current study points to the involvement of the hippocampus in integrating across variable

Table 5.2: Regions of activity in the interaction of scene type and Group.

		Tailarach		
Region	Voxels	*x*	*y*	*z*
R Middle Frontal	454	-41	19	48
R Middle Frontal	201	41	16	26
R Caudate and Putamen	136	15	12	11
L Middle Temporal Gyrus	126	-52	-9	-8
R Inferior Temporal Gyrus	97	45	-7	-30
L Superior Frontal Gyrus	93	54	37	93
L Anterior Cingulate	89	-7	39	2
L Caudate	71	-12	11	15
R-L Thalamus	64	11	-13	10
R Hippocampus	6	-32	22	-11

exposures in the Paired Scenes to support perceptual change and is consistent with recent findings that the hippocampus is involved in integrating overlapping memories during encoding (Shohamy & Wagner, 2008).

Activity in the ITG and fusiform ROI exhibited the same pattern as the hippocampus: the interaction of scene type by Group was powered by increased activity when Perceivers viewed Paired relative to Target Scenes (PSC: $t(8)$ = -2.10, p = 0.07; betas: $t(8)$ = -2.72, p = 0.026) and no response for either scene in the Non-Perceivers. Both the ITG and the FFG have been implicated in object perception in humans and non-human primates (ITG: Haxby et al., 2001; Logothetis & Sheinberg, 1996; FFG responds preferentially to faces: Kanwisher, McDermott, & Chun, 1997; non-face expert objects: Gauthier et al., 1999). Our behavioral results established that variable exposure to the novel object in the Paired Scenes catalyzes perceptual change. Differential activity of these ventral visual regions, along with the hippocampus, to the Paired Scenes in Perceivers suggests that perceptual change is supported by the combination of the sophisticated object-processing and a learning system binding across individual and

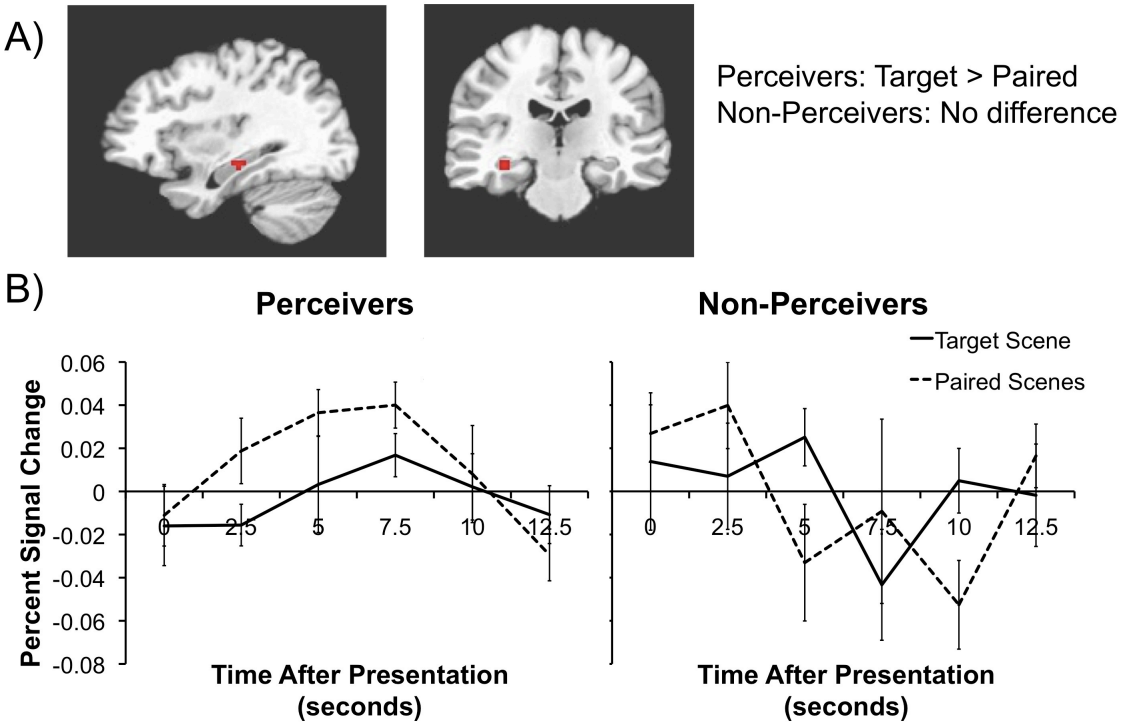


Figure 5.10: Response of the Right Hippocampus to the Interaction scene type and Group. Panel A presents the defined region of interest (ROI), calculated as a subset of a larger ROI encompassing the right middle temporal lobe. Specifics about this region are presented in Table 5.2. Panel B presents the time course of PSC change for presentation of scene type for Perceivers (left) and Non-Perceivers (right). The PSC time course supports the results of the mean PSC that the hippocampus responds to contextual scene presentation in the Perceivers only with no response to either scene type in the Non-Perceivers.

variable experiences.

Two regions of activation localized to the left and right caudate (Figure 5.11B) exhibited a similar pattern of response. Again, the interaction is supported by differences between Target and Paired Scenes for the Perceivers (PSC, Right: $t(8)$ = -7.33, $p < 0.001$; Left: $t(8)$ = -2.48, p = 0.038; betas, Right: $t(8)$ = -4.06, p = 0.004; Left: $t(8)$ = -3.21, p = 0.012; Figure 5.11A for right caudate) with no corresponding differences across scene types in Non-Perceivers. However, the differential response in Perceivers is driven by both a significant increase in

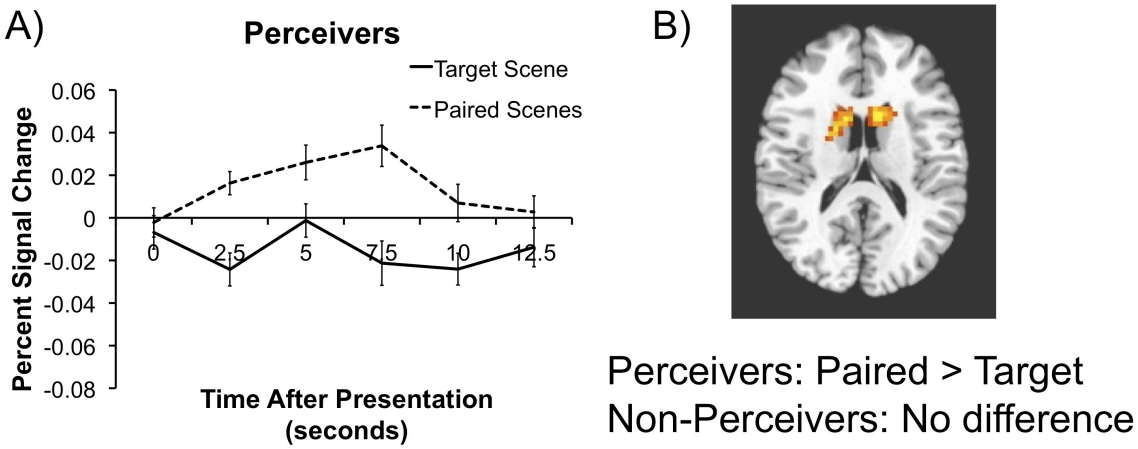


Figure 5.11: Bilateral Response of the Basal Ganglia Learning and Memory System to the Interaction of scene type and Group. Bilateral regions of activation (Panel B) localized to the caudate show the same direction differential responses to Target vs. Paired scenes by Group: differentiatal activity to Scene-type in Perceivers only. The time-course of PSC for Perceivers shows the differentiation of activity. Panel A shows PSC response plotted from the Right Caudate.

activity when viewing the Paired Scene (compared to zero: $t(8) = 9.41$, $p < 0.01$; betas: $t(8) = 2.56$, $p = 0.034$) and a significant decrease when viewing the Target Scenes (PSC: $t(8) = -5.36$, $p = 0.01$). Any significant deviation from zero in PSC is a change from baseline activity, thus a significant decrease in PSC is an indication of preferential response to the Target Scene. Activity in the basal ganglia is often considered in relation to the frequency or predictability of events (Amso et al., 2005; Redgrave, Prescott, & Gurney, 1999; Dommett et al., 2005; Schultz et al., 1997). Considering the pattern of activation in relation to frequency of scene occurrence, the Target Scene is a presented at greater frequency than the individual Paired Scenes with decreased and increased activity of the caudate respectively. Thus, Perceivers may be tracking the frequency of individual scenes. However, Non-Perceivers caudate activity does not differentiate between scene types suggesting a lack of sensitivity to the frequency of presentation in the vi-

Table 5.3: Regions of activity in the contrast of scene type. Activity for Paired Scenes is greater than for the Target Scene.

		Tailarach		
Region	Voxels	x	y	z
L Occipital and Posterior Parietal Network	1663	-14	-81	3
L Superior Parietal Gyrus	146	-19	-69	57
R Middle Occipital Gyrus	119	41	-68	-9
L Lingual Gyrus	100	-21	-96	-14
R Inferior Temporal Gyrus	92	61	-49	11

sual input.

This caudate activation is echoed in the main effect of scene type, specifically the left caudate. Also present is activation in the parahippocampus and bilateral portions of the middle occipital cortex spanning the fusiform (Table 5.3). These regions were more active for Paired relative to Target Scenes. Previous studies have linked the parahippocampal cortices to binding across surfaces in Cubist paintings (Wiesmann & Ishai, 2010) and supporting contextual processing of objects even when the context is absent (Bar & Aminoff, 2003). Consistent with these findings, this region may support perception of the Target Object across different views in both groups, while additional activity in the ventral visual stream and hippocampus, seen for Perceivers only, is necessary to change perception in the Target Scene.

Again, it is important to note that there is unequal presentation of individual scenes across scene types. Thus, we investigated whether the regions sensitive to a main effect of scene type only could reflect greater repetition suppression (RS) for Target Scenes. To this end, activity was examined over the course of exposure (7 blocks of 6 presentations). If these effects are simply being driven by unequal RS effects, one would expect a greater negative slope of activation

for the Target Scenes compared to the Paired Scenes. We did not find evidence for such an effect. We also investigated whether differential activity in the IFG, hippocampus and the right caudate are driven by Perceivers sensitivity to RS. Again, we find no evidence to suggest for RS. These results are consistent with previous studies showing that RS in object-selective regions, shown to be involved in the current task, reaches a plateau after 7 or 8 presentations (Grill-Spector, Henson, & Martin, 2006).

There is more activation in parietal regions in Perceivers relative to Non-Perceivers. The temporal regions discussed above, specifically, the bilateral middle and left superior temporal, ITG are also apparent in the main effect of Group (Figure 5.12, Table 5.4) and have been qualified above. Unique to the main effect of Group are activations in portions of the parietal lobe. These are specifically the precentral and postcentral gyri and the inferior parietal lobule (IPL). These regions exhibited greater activation in Perceivers than Non-Perceivers. This parietal activation is not sufficiently dorsal and posterior to suggest differences in visual selective attention across Groups; this is consistent with the finding of only subtle differences in eye movements across Groups. Instead these regions have been associated with novel exploration (postcentral gyrus: Binkofski et al., 1999), orientation coding (postcentral gyrus: Sakata & Kusunoki, 1992; IPL: Sakata et al., 1999; Shikata et al., 2001 and visuo-spatial transformation of objects (IPL: Harris, Benito, Ruzzoli, & Miniussi, 2008). These findings suggest that Perceivers more strongly engaged processes that help perception of an object across variable views. Additionally, activity in the IPL, part of the ventral parietal cortex, suggests a role of attentional-guided memory-retrieval in the current task. In accord with the "attention to memory" (AtoM) model (Cabeza, Ciaramelli, Olson, & Moscovitch, 2008; Cabeza et al., 2011), the IPL responds to stimuli

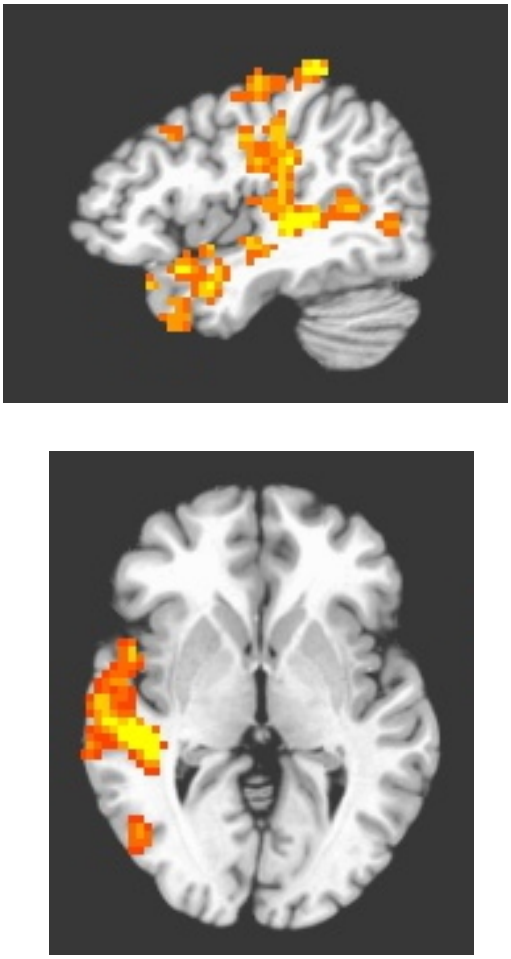


Figure 5.12: Main Effect of Group (Perceiver vs. Non-Perceivers) reveals a distributed temporal and anterior-parietal network (upper panel) including the Inferior Temporal Gyrus (lower panel). Regions in this network show greater activation for Perceivers during the task than Non-Perceivers.

that share perceptual features with a "target" both in sensory processing and memory-retrieval. Given the necessity to relate integrated representations of the Target Object to future perceptual experiences, greater attentionally-guided memory retrieval may help Perceivers to change their perception.

General differences in eye movements do not contribute to parietal activations across Groups. We examined whether general differences in looking could account for

Table 5.4: Regions of activity in the contrast of Group. Activity is greater for Perceivers than Non-Perceivers.

		Tailarach		
Region	Voxels	x	y	z
R Temporal and Anterior Parietal Network	1678	44	-31	5
L Postcentral, Precentral and Inferior Parietal Gyri	597	-50	-20	32
L Middle Frontal Gyrus	199	-34	31	40
R-L Anterior Cingulate	195	-9	-7	41
R Middle Frontal	184	39	17	33
L Superior Temporal	94	-60	-37	14
R Medial Frontal Gyrus	73	11	44	29
R Middle Temporal and Occipital Gyri	66	56	-61	-2

some of our findings of different neural engagement, notably of the parietal cortex, across Groups. First, we confirmed that there are no significant differences in general looking behavior (e.g., fixation count, total duration of fixation) across Groups for scanner participants. In order to determine whether regions related to general eye movements overlap with regions reported in the analyses above, a separate analysis considered which regions correlated with changes in fixation count and total duration of fixation (see Methods; Tables 5.5 and 5.6). Notably, for fixation count, we find activation in a number of parietal regions including the precuneus and the precentral and postcentral gyri (Figure 5.13A). While we also find Group differences in activation for precentral and postcentral gyri, the direction of the effects is opposite (Non-Perceivers $>$ Perceivers) suggesting that fixation count does not account for differences in parietal activation observed across Groups in previous analyses. For total fixation duration, we find greater activation in the superior and middle frontal gyri (Figure 5.13B) and the precuneus for Perceivers than Non-Perceivers. In sum, we find that regions typically associated with volitional eye movements support general differences in eye movements in the current task. These regions are distinct anatomically or

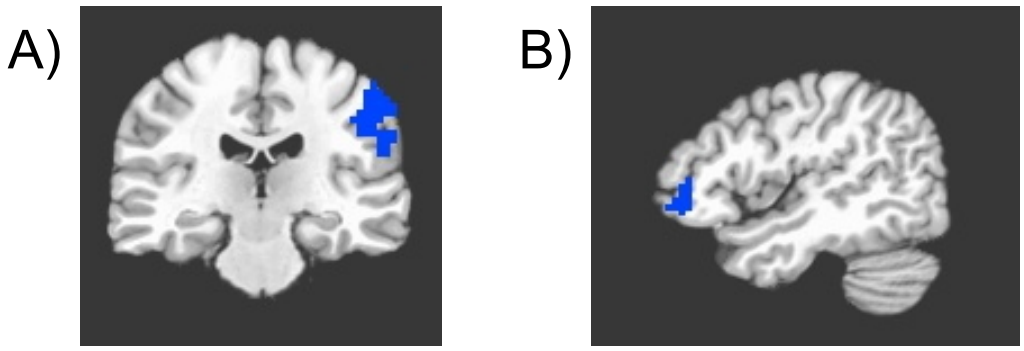


Figure 5.13: Analysis examining the regions accounting for variations in general scanning behavior. A) Differences fixation count are correlated with activity in the post- and pre-central gryi. B) Differences in total fixation duration are correlated with activity in the middle frontal gryus.

Table 5.5: Regions of activity in the analysis of fixation count. Activity is greater for Non-Perceivers than Perceivers.

		Tailarach		
Region	Voxels	x	y	z
L Post- and Pre-Central Gyrus	201	-41	-16	38
L Superior Temporal Gyrus	96	-61	-28	8
R Precentral Gyrus	58	51	-13	41
L Precuneus	51	-10	-37	47
L Posterior Cingulate	33	-4	-49	14
R Superior Temporal Gyrus	31	68	-19	8

in direction of preference from regions that differ across Perceivers and Non-Perceivers in Scene viewing. This finding confirms that differences in neural activation across Groups is not driven by general differences in scanning behavior but rather is related to sensitivity to environmental regularities in the Training condition and perceptual change.

Table 5.6: Regions of activity in the analysis of total fixation duration.

			Tailarach		
Region	Voxels	Direction of Effect	x	y	z
R Middle Frontal Gyrus	93	P greater than NP	47	38	-4
L Lingual Gyrus	59	P greater than NP	-19	-95	-14
R Superior Frontal Gyrus	36	P greater than NP	20	56	35
L Anterior Cingulate	29	NP greater than P	-8	50	-1
R Precuneus	23	P greater than NP	40	-72	36

5.3 Discussion

Using a combined eye tracking/fMRI methods approach, we examined how sampling and integration across variable experience supports changes in the domain of object perception. Adult participants were presented with a novel, occluded object in a Target Scene. Before exposure, all participants perceived this object as two disconnected parts rather than a single occluded object. We found that exposure to variable views of the object in the Training condition catalyzes perceptual change; more participants were able to perceptually complete the object after exposure in the Training condition than in the Control condition. This result establishes that, despite never seeing the novel object in its entirety, participants were able to integrate across variable and locally ambiguous experiences to arrive at a globally unambiguous percept.

We aimed to elucidate the neural and behavioral mechanisms that support the translation of variable experience into perceptual change. We found that the biggest predictor of eye movement distribution in the Target Scene was type of exposure not the presence of perceptual change: Both Perceivers and Non-Perceivers in the Training condition looked at the Object parts more than would be expected by chance while participants in the Control condition looked less

than predicted by chance-levels. We find only modest differences in eye movements between Perceivers and Non-Perceivers in the Training condition, with Perceivers looking more evenly at both Object parts which may support feature correlation (Bhatt & Quinn, 2011). Moreover, Perceivers in the Control condition did not exhibit sampling patterns like participants in the Training condition. These results indicate that sampling is sensitive to variable exposure, a catalyst for perceptual change, but appears not to directly reflect perceptual change itself in the current task. Therefore, while eye movements may serve an important function in translating environmental experience into changes in object perception, additional systems must be involved.

Considering this result in relation to early post-natal development, it is possible that regularity of experience could also drive effective sampling in infancy; this is an important area for future investigation. However, unlike in adults, infant sampling is constrained by the development of VSA (Amso & Johnson, 2006). Thus, there is a cognitive constraint on the efficacy of this mechanism in supporting early object perception.

The finding that eye movements are influenced by variable yet regular experience in the Experimental condition dovetail with recent results relating predictability with eye movements, on one hand, and an *effect* of eye movements on perceptual and cognitive processing. Starting with the latter, numerous studies have found that eye movements relate to differences in perceptual and cognitive behavior (e.g. perceptual completion in infancy Amso & Johnson, 2006; S. Johnson et al., 2004). While it has been proposed that eye movements are a window into the unfolding of cognitive processes, there is evidence that eye movements do not simply reflect but influence perceptual and cognitive processing. Grant

and Spivey (2003) found that eye movements predicted participant's arrival at a solution to a classic insight problem and, moreover, biasing participants' eye movements towards this pattern facilitated participant's ability to find the solution. (Pomplun, Ritter, & Velichkovsky, 1996) examined eye movements during perception of canonical ambiguous figures (e.g. the Necker cube) and found differences in the distribution of fixation between percepts. When the ambiguous images were modified based on these patterns of distribution (i.e. changes in contrast based on distribution of fixations), naïve participants' perception was correspondingly influenced. [1] Similarly, the current experiments provide evidence that variable yet regular exposure to a novel object can affect both the proportion of participants who change their visual object perception and the pattern of eye movements suggesting an influence of eye movements on downstream perceptual and cognitive processing.

Turning to the relation of predictability and eye movements, predictability has been found to affect fixation duration to objects in visual scenes (e.g. Wang, Hwang, & Pomplun, 2010) and also affect fixation of words during reading (for a review see Rayner, 1998). In both of these domains, increased predictability results in a decrease in fixation which might reflect greater perceptual efficiency. Indeed, increased efficiency of eye movements, such as sampling distribution, increased fixation of relevant regions and amplitude of saccades, has been found as a result of greater familiarity with visual stimuli (e.g. faces, Heisz & Shore, 2008) and with increased expertise (e.g. chess experts Charness, Reingold, Pomplun, & Stampe, 2001). These findings have suggested that eye movements can

[1]The Target Scene in the current study could be considered a type of ambiguous figure in that there easily are multiple interpretations. However, unlike figures employed by Pomplun et al. (1996) and classically called "ambiguous," we do not find that the perception of the Target Scene switches back and forth between a disconnected and a completed percept, as determined by debriefing. Instead, there appears to be a single switch in perception over the course of the experiment with many participants unaware that their perception changed.

reflect internal representations. Interestingly, we find an *increased* fixation duration to the Target Object during exposure with greater predictability (i.e. during the Experimental condition). Considering the current results in relation to this previous research, increased fixations of the Target Object could be evidence for greater focus on the relevant regions of the visual scene similar to the increased and earlier fixation of relevant chess pieces by chess experts (Charness et al., 2001). According to this explanation, exposure in the Experimental condition might also increase eye movement efficiency. However, one might expect differences in eye movement efficiency between Perceivers and Non-Perceivers, and we only find modest differences between these groups. Alternatively, increased fixation duration in this study might reflect and/or support a process of internal representational change necessary for the successful completion of the Target Object. According to this explanation, increased fixation of the Target Object might not reflect increased efficiency but be providing sensory input necessary for changes in object perception similar to the findings from Grant and Spivey (2003). While future research is needed to distinguish between these explanations, the current results broadly find that predictability can affect eye movements and that changes in exposure (in the identify of the Paired Scenes in the current study or highlighting a crucial part of a visual scene in Grant & Spivey, 2003) can affect eye movements and in turn affect downstream cognitive processes such as object perception and insight problem solving, respectively.

Neuroimaging results shed light on the additional systems supporting perceptual change. Broadly, we found evidence for that learning and memory systems, specifically the MTL, work in tandem with perceptual and attentional systems to support changes in perception. The hippocampus responded preferentially to Paired Scenes in Perceivers but does not respond to either scene in Non-

Perceivers suggesting that this learning system is involved in binding across the variable experiences with the novel object. These findings are consistent with an emerging view of the computational function of the hippocampus as integrating overlapping and related experiences at the point of encoding (Shohamy & Wagner, 2008). In the domain of object processing, computational models have pointed to a role of integrative encoding in the MTL in the production of invariant object representations (Deco & Rolls, 2004). Current results lend some support to this view: specifically that invariant object representations might be the result of hippocampal binding across variable views of an object. More broadly, these findings support the theoretical view that MTL systems are involved in visual perception and specifically, in leveraging experience to support veridical perception.

We also found evidence for the involvement of object perception, indicated by preferential activity in the fusiform and ITG. Interestingly, we do not find that these regions are preferentially active during presentation of the Target Scene, where overt changes in object perception take place, but during the Paired Scenes which provide the experiential scaffolding for perceptual change. This result suggests that these regions have a role in buttressing perceptual change. Moreover, the parallel pattern activity of these ventral visual regions and the hippocampus suggest that they work in tandem to this end.

Finally, we found that Perceivers preferentially engage regions of the parietal cortex throughout the task. Previous studies have implicated some of these regions in cognitive processes involved in relating visual experience across changing views (changes in orientation, mental rotation etc.) suggesting that perceptual processes, in addition to activity of learning systems, are essential to trans-

late experience into changes in internal representations.

It is important to consider whether the current findings are simply a result of general differences scanning behavior across participants. To this end, a number of measures of scanning behavior were examined (e.g. number of fixations, duration of fixations). We do not find consistent differences in general scanning behavior between Groups or across Conditions. We also considered which neural regions are related to Group differences in scanning behavior. This analysis revealed a set of regions that is mutually exclusive from those relating perceptual change to effective sampling. Thus, current findings are not clearly attributable to general differences in scanning behavior but more likely reflect differences in information gathering.

Recent work in developmental psychology has focused on the role of environmental regularity or statistics in shaping perception and cognition (e.g., Romberg & Saffran, 2010; Fiser & Aslin, 2002). Specifically, it is believed that prior experience with statistical information can support effective categorization and segmentation of perceptual input. Our results suggest one mechanism by which this process can proceed: The integration of variable yet regular experiences creates robust and invariant internal representations of sensory input that can shape future perception. Current findings suggest that this process is dependent on the hippocampus and effective sampling. These systems are both available to infants early in post-natal development: Behavioral research has established that the hippocampus is available for memory formation early in infancy (e.g. Little, Lipsitt, & Rovee-Collier, 1984); as reviewed in the introduction, sampling develops during the first months of post-natal development. Thus, the mechanism elucidated in the current study could help support per-

ceptual and cognitive development early in post-natal development as well as across the lifespan.

5.4 Materials and Methods

Participants.

Sixty-one healthy volunteers (37 female, *M* age = 21.39 years, *SD* = 2.8, Ethnicity: 6 Hispanic, 55 Non-Hispanic, Race: 38 white, 16 Asian, 2 African-American, 3 mixed race, 1 Hispanic, 1 other: Mestizo) completed this study. We acquired behavioral and fMRI data in 21 participants (14 female, *M* age= 21.5 years, *SD* = 2.6, all right handed). Participants had no history of neurological or psychiatric disorders, were born full-term with no major birth complications, and had normal or corrected-to-normal vision. Participants were recruited using flyers and the departmental undergraduate subject pool and were compensated for their time and travel with course credit or money. Written informed consent, in accordance with the policies of the Institutional Review Board at Brown University, was obtained from all volunteers prior to participation in the study.

Eye Tracking Apparatus and Pre-Processing.

Eye position was tracked using an SMI system (Sensorimotor Instruments Needham, MA) sampling at 60Hz and native iView software. For behavioral subjects, an SMI RED system was used with a 5-point calibration, 4-point validation routine. Average error was 0.47° and 0.50° in the x and y co-ordinates, respectively. For neuroimaging participants, an SMI iView X MRI-LR infrared eye

tracker was used with a 9-point calibration, 4-point validation routine. Average error for participants in the scanner was 1.04° and 1.72° in x and y co-ordinates, respectively.

Scenes were presented using SMI Experiment Center (2.5, 3.0) on a white screen measuring 24.5° x 12.5° of visual angle. For behavioral participants, stimuli were presented on a 22'' monitor. Tracking distance for the SMI system was approximately 70 cm. Inside the scanner, stimuli were presented by rear-projection onto a screen viewed through a mirror box mounted above the head-coil. Visual angle was calculated from the screen to the position of mirror within the magnet.

We used software specific to the SMI (BeGaze 3.0) to identify and extract fixations the four AOIs for the Target Scene. Fixations were defined with a maximum dispersion of 100 pixels and a minimum duration of 80 ms (native settings). We calculated proportion of fixation durations for each AOI per trial as the total duration of all fixations per AOI / the total duration of fixations for all AOIs. To consider eye movements as a function of exposure, data were binned across every seven presentations of the Target Scene, equaling six Training Intervals.

Stimuli and Design.

Stimuli were two-dimensional, black and white line drawings depicting visual scenes (Figure 5.1) with five to seven surfaces. In the Target Scene, the Target Object was occluded in such a way that it is unclear whether it is two separate objects (disconnected) or a single object behind another object (completed;

Figure 5.2). Exposure conditions provided equal experience with the same Target Scene but were differentiated by the identity of the Paired Scenes. In the Training condition, Paired Scenes included additional but varying views of the Target Object. Even though the object continued to be occluded, there was no ambiguity as to whether there is a single object or two separate objects as in the Target Scene. In the Control condition (tested only behaviorally), the Paired Scenes had the same number of objects or surfaces but did not include the Target Object (Figure 5.1, left and right panels respectively).

5.4.1 Procedure

Pre- and Post-Training Tests.

For the pre-training test, all participants colored the Target Scene and Paired Scenes from both conditions (8). The experimenter colored an example scene, which was not subsequently used in the study, and explained that the purpose of the study was to understand the participants visual perception. Participants were told that the scenes were intentionally designed to be unfamiliar and abstract and to color surfaces that comprised the same object the same color. Only participants who colored the object as disconnected in the pre-test were enrolled into the experiment. An additional 25 participants were found to have a completed percept of the Target Object and were subsequently excluded.

In order to ensure their appropriateness for the study before scheduling a scan, the pre-test for neuroimaging participants was conducted in a separate session. The scanning session was scheduled as soon as possible (*M* difference = 9.62, *SD* = 7.53 days). Fourteen additional participants were found to have a

completed percept during pre-test coloring and were not enrolled in this study's imaging session but completed another study in the lab. Two enrolled participants were found to have a completed percept at the beginning of the scan, as indicated by their online behavioral responses before the first block of exposure (see Task Procedure). Their data were excluded from subsequent analyses. For the post-test, participants colored only the visual scenes for their exposure condition (4).

Task Procedure.

Participants received three blocks of passive exposure to the visual scenes. During each block, participants were presented with sequential alternating Target and Paired Scenes. Each scene type was presented 14 times per block; Paired Scene order was randomly determined with equal frequency. In the lab, scenes were presented for three seconds with a three second inter-stimulus interval (ISI) during which a blank, white screen was presented. In the scanner, this ISI lasted 12 seconds in order to accommodate the time course of the BOLD response.

We gathered two measures of perceptual change over the course of the task, in addition to the eye movements. Before and after each exposure block, participants were asked to report the number of objects in each scene, without time limit, as an indirect index of whether they perceive the Target Object as complete or disconnected. Participants in the lab indicated the number of objects using the keyboard and to pressed 'SPACE' to proceed to the next scene. Participants inside the scanner verbalized their response to the experimenter. This measure, in additional to verbal debriefing, was used to identify two neuroimaging par-

ticipants who changed their percept from pre-test (disconnected) to completed prior to the first exposure block.

After the final training block, participants viewed visual scenes with numbers on each surface and were asked to use the numbers to describe how they viewed the scene (e.g. 1 and 2 are part of the same object which is behind 3, and 4 is the background). The purpose of this task was to test correspondence between verbal report of perception during the experiment with participants post-exposure coloring.

5.4.2 fMRI Data Acquisition

Images were acquired using a 3 Tesla Siemens Trio MRI scanner (Siemens, Erlangen, Germany). A 3D localizer was run (AAScout) to position the slices for the remainder of the sequences. This localizer was rotated to ensure whole brain coverage. All other images were collected in the same oblique angle. A high resolution anatomical image (MultiEcho MPRAGE: 1.20 mm isotropic voxel size, TR = 2200 ms, TI = 1100 ms, Flip angle = 7 °, 4× acceleration, 144 slices, Bandwidth = 651 Hz/Px) was collected for 3D localization and morphometric analyses.

A slow event-related design was used for functional imaging runs allowing for a direct mapping of the hemodynamic response onto each visual presentation. Echoplanar imaging (EPI) was used to measure the BOLD signal as an indication of cerebral brain activation during three blocks of exposure. EPI images were aligned to the whole brain MPRAGE anatomical image (TR = 3000 ms, TE = 28 ms, flip angle = 90°). Forty-two oblique-slices were collected of 3

mm thickness and 0 mm gap (64 × 64 mm in-plane resolution) were collected for 168 repetitions (including 2 discarded acquisitions at the onset of each of the three runs). Two participants had fewer than the 508 total repetitions collected. Due to an accidental squeezing of the emergency squeeze ball, one participant had the last 53 repetitions from the 2nd run not collected. Due to a scanner error, another participant had only 102 repetitions from the final run collected. However, both participants received the same stimulus exposure, and eye tracking measures continued to be collected.

After EPI sequences, we acquired anatomical T2-SPACE images (high resolution turbo-spin-echo with high sampling efficiency, 1.20 mm isotropic voxel size, TR = 2800 s, TE = 327 ms, 144 slices, Bandwidth = 651 Hz/Px) at locations identical to the functional image for localization purposes. Finally, functional images (T2* BOLD) were collected while participants were in a resting-state (3.0 mm isotropic voxel size, TR = 3.0 s, TE = 30 ms, Flip angle = 85°, 47 transverse slices aligned approximately to the AC-PC plane, no skip, no dummy scans, fat saturation on). All lights were turned off. Participants were instructed to simply let their minds wander and rest but to keep their eyes open and to stay awake during the scan.

Image Processing and Analysis.

Functional imaging data were processed and analyzed with the Analysis of Functional NeuroImages (AFNI) software package (Cox, 1996). EPI and anatomical images were deobliqued. Images then underwent (1) registration to the first image volume, (2) alignment to the high-resolution anatomical data set (MP-RAGE), and (3) smoothing with an isotropic 6.0 mm Gaussian kernel.

Two types of models were fit for each subject. The first model included regressors for the Target and Paired Scenes (collapsing across the three Paired Scenes). A second amplitude-modulated (AM) model was included to examine the effects of differences in eye movements on the BOLD response. The model was run twice: Separate regressors convolved the Target Scene event amplitude with 1) total duration of fixation and 2) fixation count. Both measures of eye movements were considered per trial and normalized to the average values for each subject.

Time series were normalized to percent signal change (PSC) by dividing signal intensity at each time point by the mean intensity for that voxel for that run and multiplying the result by 100. General linear modeling (GLM) was performed to fit the PSC time courses to each regressor. Linear and quadratic trends were modeled in each voxel to control for correlated drift. Motion regressors, calculated during preprocessing in three dimensions (roll, pitch and yaw) were also included.

Two separate group level analyses were run for the two model types. A linear mixed effects (LME) model included factors for subject (random effect), scene type (Target vs. Paired) and Group (Perceiver vs. Non-Perceiver) and was run within AFNI using functions from the R software package (http://www.R-project.com, Vienna, Austria, 2005). Correction for multiple comparisons was applied at the cluster level following Monte Carlo simulations conducted in the AlphaSim within AFNI. This calculation determines the probability of obtaining a false positive for the 3D image using individual voxel probability threshold in combination with a cluster size threshold. Spatial correlation between voxels was assumed. For the prescription used (64 x 64 voxel matrix, 42 slices, 3.0 mm^3

voxels) and preprocessing techniques (Gaussian filter applied of 6.0 mm), a 1000 iteration Monte Carlo simulation was run. The simulation revealed that when probability threshold is set at $p = 0.05$, 63 contiguous voxels are required to correct for false positives to $p < 0.05$. Follow-up tests determined the direction of results for LME interaction using beta weights for the region from the GLM and mean PSC (averaged across the 12.5 seconds subsequent to Target Scene presentation). For the interaction of scene type and Group, the region encompassing the right middle temporal lobe also spanned the right hippocampus. In order to consider the individual contributions of hippocampus in isolation, a separate mask was created to cover the right hippocampus only. This mask was a sphere with its origin at co-ordinates -32, 22, -11 with a radius of 3.5 mm (Figure 5.10A).

The second group analysis was a t-test (3dTtest) contrasting regions of activity correlated with the two measures of general eye movements in the Target Scene by Group. AlphaSim was run, as described above, but including only the voxels of the standard brain used in all analyses, supplied by AFNI (TT_N27), revealing that 23 contiguous voxels must reach $p = 0.05$ to correct for multiple comparisons at a level of $p < 0.05$.

CHAPTER 6

CONCLUSIONS

In recent history, there have been pronounced fluctuations in the emphasis on behavioral change or learning in the study of perception and cognition. These swings in the interest are no better illustrated than in the first words of E.J. Gibson's seminal work, *Principals of Perceptual Learning and Development*. This book was written in 1969 and begins by examining the orientation of the field.

> If anyone acquainted with the psychological literature of the last fifty years were asked what problem was of most absorbing interest to American psychologists.. he would surely answer "Learning" (E. J. Gibson, 1969, p. 1).

If the same question were asked today, the answer would certainly not be learning. Indeed, the answers that we find in contemporary analogues to E. J. Gibson's book are "the organization of the mind" or "the nature of mental representation" (Fodor, 1983; Spivey, 2007, respectively).

Psychology's interest shifted away from interest in learning or behavioral change to the internal processes of the mind began with during the Cognitive Revolution of the 1950s and 1960s. This revolution rejected the tenets of Behaviorism which asserted that the mind is best studied exclusively in its behavioral instantiation and eschewed the study of internal physiological processes in their contribution to mental processes. Behaviorists focused on change in behavior based on learning paradigms such as classical or Pavlovian conditioning and viewed mental function as highly malleable. Indeed, as the quote by E. J. Gibson exemplifies, behavioral change or learning was viewed as the most im-

portant aspect of mental life to study prior to the Cognitive Revolution. The Cognitive Revolution rejected the notion that the internal mind was a "black box" and focused on the internal structure of mental processes (e.g., Broadbent, 1958; Chomsky, 1959; G. Miller, 1956; Neisser, 1967). The rejection of Behaviorism led to a de-emphasis on everything connected to the paradigm, including learning. After the Cognitive Revolution, the study of the mind sought to carve cognition at its joints, and the processes that support experience-based change (e.g., long-term memory) were conceptualized as separate from other cognitive functions such as perception. Additionally, behavioral change has been viewed as distinct across stages of development: short- and long-term memory in adults and the process of cognitive development in infancy and childhood. Some areas of Cognitive Science have even asserted that aspects of cognition rely upon built-in or innate structures (e.g. language, Chomsky, 1967; Pinker, 1995). The view that cognition is supported by innately supplied structure or knowledge will be referred to as a nativist approach or *nativism*.

This dissertation is part of an alternate view of the mechanisms of behavioral change, a view that merges the Behaviorist emphasis on the malleability of behavior through experience and the Cognitivist emphasis on internal cognitive processes based on representations. This hybrid approach is found in many contemporary theories of development including Connectionism (e.g. Elman et al., 1996; Thelen & Bates, 2003), Bayesian inference (e.g. Gopnik & Tenenbaum, 2007; Tenenbaum, Kemp, Griffiths, & Goodman, 2011), Neuroconstructivism (Mareschal et al., 2007) and Neo-Constructivism (S. Johnson, 2010). These *empirical* approaches postulate that cognitive development is supported in large part by experience and not on innate cognitive structures and thus can be considered in opposition to nativism. Consistent with these contemporary empirical

approaches to cognitive development, the current work emphasizes the importance of experience with environmental structure in the development of internal cognitive processes and the representations upon which they act.

Specifically, this dissertation proposes that experience-based behavioral change can arise through an interaction of perceptual (auditory/visual) and learning and memory processes during experience with novel statistical regularities in sensory input. Cognitive approaches have postulated that the processes supporting experience-based change, such as long-term memory and cognitive development, are separate from other cognitive functions like perception. Consistent with this view, statistical learning has been conceptualized as a domain-general, amodal mechanism, as reviewed in Section 1.5.1. However, the research presented in Chapters 2 and 3 argues that perceptual processes can both support and constrain what is learned from statistical regularities, and structure present in sensory input can shape perception, as demonstrated in Chapter 5. Thus, perceptual and learning and memory systems appear to mutually influence each other as a result of experiences with novel statistical regularity.

This dissertation also argues that perceptual and learning and memory processes can support experience-based change throughout the lifespan. In addition to conceptualizing the processes that support experience-based change as separate from perception, these processes have also been viewed to be distinct across stages of development: short- and long-term memory in adults and the process of cognitive development in infancy and childhood. As reviewed in Sections 1.3.1 and 1.5.1, statistical learning has been demonstrated in both infants and adults. Indeed, previous studies have argued for invariance in the

mechanisms of statistical learning (Kirkham et al., 2002; Saffran et al., 1997). This dissertation builds on the view that statistical learning mechanisms are continuously available but argues that at all stages of development, this mechanism operates through a dynamic interaction of perceptual and learning and memory systems. Such a dynamic interaction will result in changes in learning based on perceptual processes. Chapters 3 and 4 present evidence for this view: Experiments in these chapters examine learning across auditory and visual modalities in adult and infant learners, respectively. Consistent with previous studies, results in Chapter 3 find that adults tend to have superior statistical segmentation abilities in the auditory modality (without temporal disruptions and at a fast rate of presentation as investigated in Section 3.2). Extending this work with adult learners, Chapter 4 presents the first comparison of statistical learning across perceptual modalities in infant learners. We find no evidence for auditory superiority in infant learners. In fact, we find evidence for an earlier availability of visual SL in infancy. Comparisons of learning across perceptual modalities in infancy suggests that 1) statistical learning is also affected by perceptual processes across the lifespan, and 2) there is a different influence of perceptual processing in infancy than in adulthood.

In addition to finding that learning from statistical regularities is influenced by perceptual processes in both infants and adults, this dissertation also presents evidence that activity of learning and memory systems, in response to experience with regularities in sensory input, can influence perception suggesting a dynamic interrelationship between these systems. Chapter 5 examines how variable yet regular experience with a novel object affects perception of that object. This study used a combined eye-tracking/fMRI methods approach to examine the nature of changes in object perception as a result of this expe-

rience. Neural activity in the medial temporal lobe and specifically the hippocampus was found to correlate with changes in object perception as a result of experience. While this dissertation does not present empirical findings that experience with structure in perceptual input directly affects perception in infancy, Section 5.3 argues that the mechanisms found to support changes in object perception as a result of structured experience could be operational in the first post-natal year. While future research is necessary to directly test these findings in infant learners, the work presented in this dissertation suggests that a dynamic interrelationship between learning and memory systems and perceptual systems can support experience-based changes across the lifespan.

This final chapter discusses the findings presented in this dissertation in relation to the broader goal of Cognitive Science: to understand the human mind. Two separate sections discuss the implications for the role of perception in learning from statistical regularities and the implications of learning and memory and statistical regularities influencing perception. Finally, specific implications are discussed for the fields of Developmental Psychology, Cognitive Science and Cognitive Psychology and Cognitive Neuroscience.

6.1 Statistical Learning Is Influenced by Perceptual Processing

This dissertation presents evidence that perceptual processing affects statistical learning. These findings are particularly relevant because: 1) it contributes to an understanding of the cognitive and neural mechanisms supporting statistical learning, and 2) it addresses the question of how statistical learning unfolds outside of the laboratory or *in the wild*. This section discusses how a perceptual-

influence on SL can provide insight into both of these open questions. Section 6.1.1 expands this discussion to the effect of perceptual variability on statistical learning.

Following from the first point, a standard view of the mechanisms of statistical learning, as reviewed in Section 1.5.1, asserts that learning is affected by relevant statistical information in sensory input (e.g., transitional probabilities). It follows from this view that sensory input with different perceptual character but identical statistical information would result in the same learning outcomes. This view of statistical learning suggests that the underlying mechanisms are singular and amodal or abstract and that statistical learning operates based on unitized perceptual output (e.g. object identity, phonemes, or syllables).

However, if statistical learning is affected by additional perceptual information, this challenges the standard view of the mechanisms of statistical learning. As outlined in Section 4.1, sensory input can be conceived of as containing both relevant statistical information and additional non-statistical perceptual information. Perceptual information could be considered differences in modality, whether statistical information is presented as part of auditory vs. visual stimuli as in Chapter 4[1] but can also be the type of stimulus within a perceptual modality such as whether visual stimuli are geometric shapes or faces. These factors have been referred to as "statistical information" and "perceptual infor-

[1]In this case, it is important to differentiate the contributions of different sensory transduction processes from differences in perceptual processes. Sensory processes are largely affected by the sensori-motor nature of the particular sensory system; for example, visual sensory processes are constrained by the organization of the retina and thus operate based on a retinotopic organization. However, some character of the sensory organization is inherited by perceptual processes even though the sensori-motor constraints typically weaker in perceptual processes. (see Cleland, 2010, for an insightful discussion on this topic) Thus, it is difficult to dissociate relative contributions of sensory transduction and perceptual processes when comparing across sensory modalities. The approach taken in the current work is to control for basic differences in discriminability across sensory modalities. See Chapter 4 for a deeper discussion on this point.

mation," respectively.

There is evidence that statistical learning is sensitive to non-statistical information. Chapter 3 found evidence that statistical learning is not equivalent across perceptual modalities. This study examined learning when identical statistical information is presented as instantiated in both visual and auditory modalities and how perceptual conditions, influenced learning outcomes. Perceptual conditions were influenced through changes in presentation rate. Results demonstrated that learning is significantly affected across modalities and perceptual conditions. Specifically, statistical learning proceeds for a given modality only in amenable perceptual conditions. Additionally, the experiments reported in Chapter 3 also examined the influence of selective attention. Recent research has suggested that statistical learning may also be affected by selective attention, as reviewed in Section 1.5.1. Baker et al. (2004) and Pacton and Perruchet (2008) suggested that selective attention may be the mechanism by which perceptual conditions, such as perceptual grouping, influence statistical learning. However, we found no evidence that selective attention could compensate for poor perceptual conditions. In fact, the results suggest that the effects of selective attention on statistical learning are also affected by perceptual conditions: There was no increase in statistical learning from attended stimuli in poor perceptual conditions. Overall, these results suggest that selective attention and perceptual processes each influence statistical learning.

Overall, the results from Chapter 3 suggest that statistical learning is mediated by mechanisms that are sensitive to the perceptual nature of the input in addition to its statistical structure (e.g., transitional probabilities and frequencies). Neuroimaging evidence has also suggested just such a scenario: Turk-

Browne et al. (2009) report increased visual cortex activity, specifically in visual object region LOC, during the observation of learnable visual sequences. Taken together with the behavioral results supporting differences in learning across perceptual modalities, this pattern of activity suggests that the visual cortex is likely involved in learning the underlying structure from visual statistics (in addition to results from Chapter 3 see Conway & Christiansen, 2005, 2009; Robinson & Sloutsky, 2007; Saffran, 2002).

In light of these findings, Chapter 3 summarizes three possible architectures for a cognitive mechanism supporting SL (or ISL as referred to in Chapter 3). Schematics of these mechanisms are presented in Figure 3.6. The aim of this work is not to conclusively support one type of mechanism for statistical learning but to further elucidate the importance of perceptual processing in learning from statistical regularities. However, as asserted in Chapter 3, both the modality-specific and the embodied mechanisms more readily support differential learning across the visual and auditory modalities, while the standard view, which suggests an amodal or abstract mechanism, does not readily fit with the findings of modality-specifity such as those in Chapter 3.

Findings of modality-specific learning also provide a entrée to understanding how statistical learning proceeds outside the laboratory. This section now turns to the second point above, that understanding how perceptual processing affects statistical learning is important for understanding how statistical learning unfolds outside of the laboratory or *in the wild*. The field has expanded evidence of statistical learning to many different types of statistical information present in sensory input. In these experiments, the only learnable statistical information presented to the observer is along a single, relevant dimension. For

example, when examining statistical learning of sequences, there is no variability in the presentation of the objects (i.e., no distributions in sensory information, as in Kirkham et al., 2002; Saffran et al., 1996). Similarly, studies examining learning from distributions of sensory information present this information in non-predictable sequences (i.e., no higher-order statistical information is present, as in Maye et al., 2002; Yoshida et al., 2010). Outside of the laboratory, statistical information is typically present along multiple dimensions of sensory input simultaneously. Thus, statistical learning experiments that examine learning along a single dimension do not provide a clear idea of how learning proceeds outside the lab where learning is a multidimensional problem. This issue is discussed at length in Chapter 2 and the paper presented in Appendix D. The finding that perceptual processing affects statistical learning may make the problem of understanding how statistical learning operates in the wild more tractable.

One of the major difficulties in understanding how statistical learning mechanisms operate outside the laboratory is the lack of constraints on learning and the ubiquity of statistical information in sensory input received from the environment. Indeed, a domain-general, amodal learning mechanism posits that learning depends only upon the presence of statistical information, and thus this type of mechanism has no obvious constraints.

Considering a mechanism where statistical learning is affected by perceptual processing provides an initial view of how learning can be constrained to a subset of the statistical information. Chapter 2 examines how perceptual processing can bias the multidimensional statistical learning task of language acquisition. Specifically, the relevant statistical information was defined across groups or

categories of novel sounds and participants were tested as to whether they segmented the stream based on this higher-order statistical information. In this way, the statistical learning paradigm was similar to the studies of Kirkham et al. (2002); Saffran et al. (1996) and the experiments presented in Chapter 3. However, these experiments presented exact repetitions of objects in different orders to produce higher-order statistical regularities. Diverging from this methodology, the experiments in Chapter 2 asked participants to segment pairs from the auditory stream based on variable presentations or multiple exemplars from novel sound categories. Thus, participants must both resolve the perceptual variability to determine which sounds conform to the same category as well as use the information defined across categories, not single sounds, to segment the acoustic stream. The inclusion of perceptual variability across successive experiences allowed for the investigation of how perception of the sounds affected learning and specifically, whether the perceptual organization of the sounds supported the categories used to define the higher-order statistical information. Two broad theoretical possibilities were contrasted: 1) statistical learning is primarily "top-down" – that is, statistics defined across sound categories can be optimally learned, regardless of the perceptual organization of these sounds; 2) statistical learning is primarily "bottom-up" – that is, the perceptual organization at lower levels is relied upon to define statistical information at higher levels. The results are more consistent with the latter view: Perceptual organization constrained what statistical information was learned (see Chapter 2 for an indepth discussion).

Broadly, the results reported in Chapter 2 suggest some constraints on the statistical learner: If higher-order statistical information is not supported by lower level perceptual organization, then the statistical information is not read-

ily learned from, and the corresponding environmental structure is not uncovered. This finding has important implications for the conceptualization of how statistical learning contributes to language acquisition, as discussed at length in that chapter. Relevant to the current point that considering how perception affects statistical learning provides some insight into how learning can proceed in the wild, the differences in perceptual abilities, such as speech categorization, will provide different access to statistical information present in the speech stream. Thus, any change speech categorization abilities could potentially reveal new statistical information to the observer. Thus, perceptual processing can be said to apply some constraints on the statistical learner and correspondingly some indication of what would be learned from the statistically-rich sensory input received outside the lab.

There are similar implications for the results presented in Chapter 3: What statistical learning outcomes are constrained by the perceptual modality and the corresponding perceptual conditions. Specifically, at slow rates of presentation, visual statistical regularities were learned from but not auditory statistical regularities; however, at fast rates of presentation, the opposite pattern was found, where auditory statistical regularities were learned from but not visual statistical regularities. These findings suggest that the same auditory and visual statistical regularities are learned from under different perceptual conditions.

This perceptual constraint suggests that different aspects of the statistical information present in sensory input will be learned within different modalities. To illustrate, consider a single multisensory input endowed with statistical regularities such as the audiovisual input of a talking person. Following from the results in Chapter 3, the visual system would be more likely pick up on the sta-

tistical regularities that unfold more slowly over time (e.g., the manual gestures of the talking person) while the auditory system may pick up on the statistical regularities that occur more quickly (e.g., the rapid-fire statistical information present in fluent speech). Thus, perceptual constrains across modalities suggest that different statistical regularities from the same sensory signal will be learned across modalities, a kind of division of labor.

Moreover, we find evidence that differences in statistical learning across perceptual modalities can change over developmental time. Chapter 4 provides evidence that visual statistical learning is possibly more robust than auditory statistical learning. In fact, we see evidence for more robust visual learning in infants younger than 9.3 months and no evidence for auditory superiority in older infants. This result is surprising given the typical pattern of auditory superiority in adult learning. These results demonstrate that there is not a static relationship between perceptual processing and statistical learning with a different pattern of learning across modalities found in infancy and adulthood. Similar to how changes in speech categorization abilities could reveal different statistical information present in the speech stream, differences in statistical learning across sensory modalities could shift what statistical information is learned from across development.

Section 1.5.2 argued that statistical learning is grounded by the observer's niche and specifically, that the effect of an observer's niche on learning could be mediated by perceptual processing. This dissertation presents evidence in Chapter 2 that statistical learning is constrained by perceptual organization and in Chapter 3 and 4 that statistical learning is affected by the specific perceptual modality receiving the relevant sensory input across the lifespan. This evidence

that statistical learning is affected by perceptual processing provides insight into the nature of the mechanisms supporting statistical learning, and specifically that an amodal view of statistical learning is not adequate to capture the differences in learning seen across perceptual conditions.

6.1.1 The Effects of Perceptual Variability on Statistical Learning

The previous section argued that understanding how perception affects statistical learning will provide insight into the mechanisms underlying statistical learning and how these mechanisms operate in the wild to support experience-based developmental change. Building on the argument that perceptual processing affects statistical learning, this section asserts that *perceptual variability* can also affect statistical learning. Learning from experience with statistical regularities cannot occur as a result of a single event or experience. Instead, learning from statistical information as conceived by the field of statistical learning (e.g., transitional probabilities, distributions of frequency) must occur as a result of numerous individual experiences (see Section 1.4.2). However, most statistical learning studies, focused on how statistical information alone affects learning, have typically employed repetitions of objects and do not consider how variability across these multiple experiences affects learning.

Sensory input varies widely across multiple experiences of even the same object or utterance. Taking the example of language, it is well known that there exists a large amount of acoustic variability across the multiple utterances of any given linguistic unit, even within the productions of a single speaker. Chapter 2

reviews the ways in which speech variability necessitates the development of speech categorization to map variable utterances to their functional categories. Variability of sensory input is even possible within the same objective sensory input. Chapter 5 emphasizes that visual sampling or eye movements affect sensory input and the patterns of eye movement are related to perceptual abilities in infancy and appear to co-develop with selective attention (e.g. Amso et al., 2005; S. Johnson et al., 2004).

Despite the ubiquity of variability across experiences, little is known about how statistical learning is affected by perceptual variability. As reviewed in the previous section, most statistical learning experiments present statistical information along a single, relevant dimension of the sensory input. For example, using exact repetitions of utterances. While additional statistical information could be included by adding an informative distribution to the utterances, additional perceptual or non-statistical information could be added by permitting variability across experiences with an utterance. The inclusion of perceptual variability would bring statistical learning paradigms closer to the sensory experience received outside the laboratory, but few experiments have investigated learning in paradigms that include non-statistical, perceptual variability.

An understanding of how perceptual variability affects statistical learning is of increased importance given the current demonstrations that perceptual processing affects learning. If statistical learning were supported by a domain-general, abstract mechanism focused on statistical information only, perceptual variability would not be relevant. Following from an abstract view of statistical, the role of perception in statistical learning could simply be the detection of the occurrence of an object or sensory experience or the discrimination of a given

object from the set of objects used in the study. According to this view, perceptual variability could affect statistical learning in so far as perceptual variability affects discrimination of individual objects. However, if perceptual variability would likely be irrelevant to statistical learning as long as the identity of the object remains equally discriminable. However, this dissertation presents evidence that perceptual processes have a much greater influence on statistical learning than proposed by standard accounts, and thus, perceptual variability could have a large effect on statistical learning.

The empirical results presented in Chapter 2 provide evidence that learning from higher-order statistical regularities can proceed despite the presence of perceptual variability. These experiments required participants to statistically segment a stream based upon four novel acoustic categories, each instantiated in six different exemplars. Because the higher-order statistical information was not reliable based upon a single exemplar or pair of exemplars, participants had to group the variable exemplars in order to uncover reliable statistical information. Overall, we found evidence that adult statistical learning is robust to a large amount of acoustic variability.

However, the way in which learning proceeds with variable exemplars results in perceptual restriction of what is learned. Participants appear to use their naïve perceptual organization of the sounds to support learning in the presence of perceptual variability instead of focusing only on the top-down optimal statistical structure of the auditory stream. Specifically, participants group the variable exemplars into three clusters and use these clusters to uncover the higher-order statistical information. The authors found that using perceptual organization to learn despite the presence of perceptual variability is both a

benefit and a detriment to learning. Participants are able to generalize their knowledge gained through statistical exposure to novel sounds that occupy the same perceptual space or clusters, suggesting that perceptual processes can provide a knowledge robust to perceptual variability. However, given that these clusters differ from the four experimenter-defined categories, the participants have suboptimal learning and poorer prediction of successive sounds. While not directly tested, it is most likely that if participants had been given a single, prototypical exemplar from each category (i.e., no perceptual variability), learning would have occurred optimally because perceptual processes would not have been needed to resolve the variability and separate overlapping acoustic categories. Similarly, if participants had been given varying exemplars for categories that are already supported by perceptual processes, learning would also have occurred optimally (e.g. Brady & Oliva, 2008; Baldwin et al., 2008). Thus, learning in the presence of perceptual variability changes the character of learning and specifically that the presence of perceptual variability may enhance the constraints placed on learning by perceptual processes.

While it seems intuitive that perceptual variability would obfuscate statistical patterns in sensory input and require additional processing to overcome variability to learn, as reported above, there is also evidence that variable experience can support more robust, behaviorally-effective learning. The learning paradigm employed in Chapter 5 also employs perceptual variability, but the presence of perceptual variability appears to support learning, while simple repetition does not. Specifically, the same Target Scene was employed in both an Experimental Condition and a Control Condition. This Target Scene contained a novel Target Object that was occluded such that it was ambiguous whether there was a single, occluded object or two separate objects. All par-

ticipants initially perceived the Target Object as two disconnected parts. The same repetition of the Target Scene was present across conditions, which provided 100% co-occurrence between these two object parts. However, in the Control Condition, very few participants changed their perception. Participants in the Experimental Condition received additional, variable exposure to the Target Object. Despite every experience being locally ambiguous, as the object was always occluded in some way, many more participants receiving this additional variable exposure changed their perception of the Target Object. This experiment suggests that variable exposure may in fact support better learning rather than obstructing learning or obfuscating statistical information.

Building on this finding, there is also evidence for a relationship between the presence of variability during exposure and the level of abstraction at which statistical information in learned. In Section 1.4.2, it was suggested that there is there is a relationship between amount of variability given during exposure and the level of abstraction required by a given learning task with tasks requiring rule-like or abstract knowledge providing a greater amount of variability during exposure (e.g. rule-learning tasks, Marcus et al., 1999). While statistical learning tasks typically display both an intermediate amount of variability and abstraction, an increased level of variability might support learning at a more abstract level of the sensory input. A recent study found evidence for category-level learning in a statistical learning task with generalization to novel exemplars (Brady & Oliva, 2008). This study also exposed participants to many different exemplar scenes for each category during a long exposure, so participants would not have been able to learn based on specific exemplars. The paper presented in Appendix D presents evidence that when observers could learn from statistical regularities based on semantic categories as well as object-identity,

learning is constrained to the less abstract, object-based regularities. However, as with the results from Brady and Oliva (2008), if the statistical information based on object-identity is not reliable, participants will learn based on semantic categories. This result suggests that statistical learning might be biased towards less abstract statistical regularities. However, the presence of perceptual variability at lower levels of abstraction during exposure can support more abstract learning.

A relationship between variability and learning of deeper or more abstract aspects of the sensory input was suggested by E. J. Gibson (1969), specifically that variability of experience could allow the observer to distinguish what aspects of the sensory signal are an integral part of a higher-order pattern or invariant. Again, rather than obfuscating statistical learning, these results suggest that perceptual variability might reveal important statistical information and specifically that variability may be necessary for observers to learn from more abstract statistical information. A link between higher variability and more abstract learning suggests another way to elucidate how learning can proceed from naturalistic sensory experience. For example, for sensory experiences supported by more abstract statistical information, greater variability of experience may be necessary for the appropriate learning to take place. The opposite may also hold that the more variability inherent in sensory input, the more abstract any patterns of learning.

However, not all variability is likely to affect statistical learning. As illustrated in J. J. Gibson and Gibson (1955), participants are able increase their ability to pick out a specific novel drawing despite being presented with perceptual variability. Relevant perceptual variability was presented along three dimen-

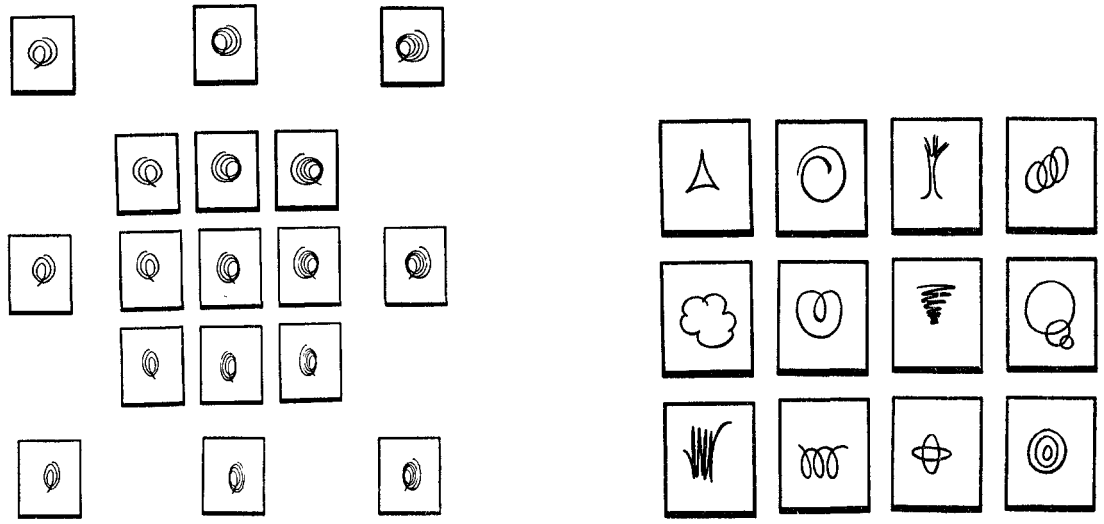


Figure 6.1: From J. J. Gibson and Gibson (1955), the left panel presents nonsense items differing along three dimensions and the right panel presents items differing along many dimensions

sions: number of circles, tightness and direction (see left panel of Figure 6.1). Without feedback, both adults and children were able to increase their ability to pick out the standard items (located in the center of the left panel of Figure 6.1). However, their ability to pick out the target item amongst perceptual variability will not increase if variability is presented along too many dimensions (see right panel of Figure 6.1). Thus, for perceptual variability to affect learning, it must be along the relevant dimensions or bear some perceptual relationship with what is being learned.

One unique contribution of this dissertation considers variability of perceptual experience in relation to learning from statistical information. Overall, these studies find that learning can be robust to large amounts of perceptual variability even without the benefit of specialized or explicit categorization abilities (see Chapter 2). In fact, the presence of perceptual variability might support invariant object representations (see Chapter 5). Finally, Appendix D presents evidence that the level of abstraction learned by an observer from sta-

tistical information might change based upon the presence of variability, and correspondingly that variability might be necessary for learning in some cases. The effect of perceptual variability on statistical learning suggests that the underlying mechanisms of statistical learning are intimately tied with perceptual processing. Additionally, the fact that perceptual variability has an impact on statistical learning (e.g., more variability results in more abstract knowledge) provides another way to understand how learning can proceed from naturalistic sensory experience. For sensory experiences supported by more abstract statistical information, greater variability of experience may be necessary for the appropriate learning to take place.

6.2 Learning and Memory Systems Can Affect Perception

The thesis of this dissertation is that perceptual systems mutually influence learning and memory systems during experience with novel statistical regularities in sensory input. The previous section discussed in depth the findings and implications in relation to how perceptual processing, and relatedly perceptual variability, can affect learning outcomes or memory. This section discusses the converse claim: that learning and memory systems can affect perception.

The broader argument in this dissertation suggests that differences in experience can support changes in perceptual processing through the response of learning and memory systems to this experience. This form of feedback loop supports dynamic change in how the observer perceives the world and in turn affects what aspects of future experience are learned from.

This dissertation argues that patterns of statistical learning across percep-

tual modalities differ in infants and adults. It is an open question as to why these developmental changes occur. However, recent work reviewed in Chapter 4 has suggested that the pattern of visual input is markedly different for infants. Through the use of head-mounted cameras, Smith et al. (2011) found that the visual input for infants 17- to 19-months-old was often dominated by single objects in succession while the adults' field of view often contained visual information from a number of objects. Thus, compared to adults, the infants' everyday visual input may be similar to the visual sequences presented in the current study: sequential presentations of single objects. One possibility is that visual statistical learning abilities in the learning paradigms employed in Chapter 4 shift as visual input shifts from infant to adult views. We also found evidence for a shift in auditory statistical learning between the ages 8 to 10 months. This shift in statistical learning abilities could be supported by significant changes in language processing abilities occurring at this age (Bates et al., 1998). While future research will be necessary in order to establish a direct link between language development and changes in auditory statistical learning as well as between visual processing and visual statistical learning, the framework presented in this dissertation suggests that differences in experience drive changes in perception and then changes in statistical learning. An integral part of this view is that differences in experience must be able to result in adaptive changes in perception.

This section discusses the effects of statistical regularities on learning and memory systems and how this influence in turn affects perceptual systems. First, an argument is presented that even though statistical learning is a passive and incidental type of learning task, statistical learning is not the result of a passive absorption of statistical information. Instead, the author argues that

statistical learning may be supported by the active process of prediction and generating prediction errors. Then, the relevant empirical study reported in Chapter 5 is reviewed and integrated with broader evidence that predictability affects activity in perceptual cortices.

6.2.1 Is Statistical Learning Active?

Statistical learning and other forms of incidental learning are behaviorally passive in that they do not require an overt, relevant behavioral response for learning to occur. However, the current section proposes that incidental learning can be considered active in so far as it may involve the process of prediction and reduction of prediction error.

It is a prominent distinction within the area of learning and memory, whether a learning task is active or passive. Tasks that are considered *active* often involve the participant engaging in behavioral responses that are relevant to the learning outcomes. For example, reinforcement learning tasks often provide feedback to a participant's actions and this feedback is necessary for learning (e.g. Shohamy, Myers, Grossman, et al., 2004). *Passive* learning tasks can often have the same computational character as active learning tasks but either do not require an overt behavioral response at all or require a behavioral response that is not relevant for what is learned. Passive learning is often also referred to as *incidental learning*. In the absence of a role for overt and relevant behavior during the learning process, learning can be seen to occur as an unplanned by-product of some other process or incidental to experience.

There is inconsistent evidence as to whether the involvement of a behav-

ioral response essentially changes the quality of the learning task or the neural systems supporting learning. There is some evidence that active and passive learning have different neural underpinnings. For example, work by Shohamy, Gluck, Myers and colleagues have found that the presence or absence of feedback affects whether the learning task is supported by basal ganglia learning and memory systems (Shohamy, Myers, Onlaor, & Gluck, 2004; Shohamy, Myers, Hopkins, Sage, & Gluck, 2009). Touching on the vast results from the animal learning and memory literature, separate neural mechanisms have been proposed for instrumental vs. classical conditioning paradigms, active and passive learning tasks respectively (e.g., Berridge, 2000). However, studies of associative or reinforcement learning in humans have found parallel involvement of the basal ganglia during learning for activity and passive versions of a reinforcement learning task (e.g., McClure, Berns, & Montague, 2003; Pagnoni, Zink, Montague, & Berns, 2002). Thus, there is some compelling evidence that the active vs. passive character of behavioral responses importantly affects the mechanisms underlying learning; however, recent work with human participants has suggested that similar neural mechanisms can support learning in both active and passive paradigms.

Considering the active-passive dichotomy presented above, statistical learning is certainly a passive learning task consistent with other forms of incidental learning. Participants in statistical learning tasks are never given a directive to learn from the patterns present in sensory input. In fact, explicit instruction to learn from sensory input affects and sometimes disrupts learning abilities (e.g. Reber, Kassin, Lewis, & Cantor, 1980). Participants often provide no responses at all during exposure to test stimuli. In paradigms where participants are asked to provide a response, the response is not relevant to the learning as

the response does not help to reveal the underlying structure of the test stimuli and is specifically designed to be orthogonal to the information to be learned (see experiments in Chapter 2 and 3 for non-relevant behavioral tasks during exposure). Thus, statistical learning tasks are unambiguously behaviorally passive.

Despite no active external change in behavior, incidental learning could be considered to be active in the sense that the brain could be learning through predicting successive sensory inputs and reducing prediction error (see Section 1.2.1). A passive view of learning from statistical regularities would involve no prediction or effort on the part of the observer to reduce their prediction error. For example, passive statistical learning of a sequence of stimuli would involve reflexive, as opposed to predictive, association where each successive stimulus simply produces an update on the associative strength, frequency, etc. for that stimulus and the previous one. An active view of statistical learning would posit that the presentation of a stimulus would result in a prediction of what the next stimulus will be, and in turn this prediction can generate an estimate of prediction error. As a result of prediction error, the observer can update their internal model of the environment. This active model of incidental learning is broadly consistent with contemporary views of learning as a reduction of prediction error (McClelland, 2002; Redgrave & Gurney, 2006; Rescorla & Wagner, 1972) and specifically with the Predictive Coding Theory of cortical function proposed by Friston (2005) which involves both prediction or predictive coding as well as prediction error. The Predictive Coding Theory of cortical responses will be discussed a more length later in this section.

One of the implications of the view of statistical learning as a passive accu-

mulation of statistical information is that more experience with statistical regularities will result in better statistical learning. However, there is some suggestion that the amount of experience is not the only factor affecting learning outcomes. First, there is evidence that the equivalent amount of exposure results in greater learning when the relevant information is selectively attended. There is also some evidence that experience without selective attention results in no statistical learning (Toro et al., 2005; Turk-Browne et al., 2005 and results presented in Chapter 3, however see Saffran et al., 1997). Second, results from Chapter 4 suggest that greater exposure might not support greater learning. The experiments in this study employed infant-controlled habituation methods which allow a greater measure of the infant's selective attention than fixed habituation. [2] In both visual and auditory conditions, non-habituated infants received greater, overtly attended exposure to statistical information (this result was significantly reliable in the auditory condition only). However, non-habituated infants do not show evidence for greater learning. In fact, both groups of non-habituated infants fail to show a statistically significant direction of preference at test, the measure for statistical learning for that experiment (see Section 4.4.5 for a discussion of these data). Again, this result entail that greater exposure does not necessarily support greater statistical learning suggesting that the passive absorption of statistical information is not the sole factor supporting statistical learning. An emerging line of evidence exploring suggest that infants' social knowledge affects statistical learning. The use of a social attention cue (e.g. eye gaze) helps infants learn in a difficult but ecologically-valid statistical learning task (Wu, Gopnik, Richardson, & Kirkham, 2011). Without a social attention cue, infants receive almost twice the exposure to the statistical information but

[2]However, infant-controlled habituation only controls for overt attention as measured by eye gaze and not other more precise attentional measures such as heart rate deceleration (Richards & Gibson, 1997).

do not show as robust learning effects (Wu & Kirkham, 2010). Providing parallel results from those presented in Chapter 4, Wu, Kirkham and colleagues find that differences the internal processes (such as selective attention) facilitated by the presence of an informative social cue result in greater learning even when infants receive much less overall exposure. Overall, these findings suggest that greater exposure to statistical information does not necessarily result in greater statistical learning. This provides some indication that a view of statistical learning as a passive accumulation of statistical information might not capture the nuances of the learning process in these tasks. Instead, internal, possibly active, processes, such as socially-based selective attention, appear to influence what is learned from experience with statistical regularities.

There is behavioral evidence that statistical learning involves processes of prediction error, a process generally believed to underlie active learning processes. Learning through the reduction of prediction error is well captured in the Rescorla-Wagner (R-W) model (Rescorla & Wagner, 1972). Famously, this model predicted two curious behavioral effects: blocking and overshadowing. [3] According to R. Miller et al. (1995), the *a priori* prediction of overshadowing and blocking by the R-W model are among the model's great successes. There is some evidence for blocking and/or overshadowing effects in statistical learning. Previous exposure to stimuli employed in statistical learning tasks affects later learning. Evidence for reduction in learning has been found when

[3] In both blocking and over-shadowing procedures, two CSs are paired with a single US. In the case of *blocking*, prior exposure of CS1 and the US prevents the conditioning of the second CS. In short, blocking is the deficit of a conditioned response to CS2 after CS1-CS2 → US in which CS1 was previously paired with the US. The former event is explained by the R-W model as the blocked CS having no associative strength. In the case of *overshadowing*, both CSs are presented with the single US from the beginning, but one CS develops a stronger CR than the other. This can be modulated by the relative intensity or salience of the CSs, for example. In this case, the prediction error is attributed to the more salient stimulus (R. Miller, Barnet, & Grahame, 1995).

stimuli have been previously presented in random temporal sequence (Catena, Scholl, Isola, & Turk-Browne, 2010) or in an unpredictive context (Jungé, Scholl, & Chun, 2007). In addition, previous learning of a statistical sequence (Gebhart et al., 2009) leads to decreased learning of a second sequence using the same stimuli. These behavioral results are consistent with predictions from the R-W model which relies upon prediction of successive stimuli for learning suggesting that statistical learning is also supported, at least in part, by predictive learning.

There is evidence that statistical learning, and incidental associative learning more broadly, includes a prediction error signal in the striatum. Striatal responses to unpredicted events have been mapped onto prediction error as conceptualized in a number of prediction-based learning models such as the R-W model. This pattern of results has been found in both passive and active reinforcement tasks in human learners (e.g., McClure et al., 2003; Pagnoni et al., 2002). Evidence of a striatal involvement in prediction error has also been found in an incidental learning task similar to statistical learning: den Ouden, Friston, Daw, McIntosh, and Stephan (2009) found that when a stimulus (an auditory tone) predicted the presentation of a second, visual stimulus, there is evidence that striatal responses represent prediction error. Thus, in passive learning tasks where the only basis for learning is the statistical regularity of successive stimuli, there is evidence for the involvement of prediction error. While this study did not specifically examine whether such responses are consistent with prediction error, Turk-Browne et al. (2009) found preferential responses of the striatum during experience with statistically-defined sequences of visual stimuli. Importantly, there was no difference in the relative frequency of individual items, another possible reason for striatum involvement in a sequence

learning task (Amso et al., 2005). Thus, there is some evidence that incidental learning broadly, and statistical learning specifically, involve prediction error as evidenced in striatal function.

Finally, recent studies have argued that perceptual cortices exhibit predictive responses as a result of experience with statistical regularities in sensory input. These studies support a theory of cortical responses called Predictive Coding. As summarized in the paper by den Ouden et al. (2009),

> Predictive coding posits a hierarchy of connected brain areas in which each level strives to attain a compromise between information about sensory inputs provided by the level below and predictions (or priors) provided by the level above. The central learning principle is to establish a good model of the world, which is achieved by changing connection strengths such that prediction errors are minimized at all levels of the hierarchy. ... Predictive coding may be a general principle of brain function in which statistical relationships in the world are monitored, even when they are .. not relevant for ongoing behavior. This would allow the brain to ignore predictable and therefore uninteresting events in the environment, thereby enhancing the saliency of unexpected events. (p. 1849)

This model is often contrasted with a feature-based view where bottom-up stimulus features drive the cortical responses and responses decline with repetition of the same stimulus features in what is called repetition suppression (e.g., Grill-Spector et al., 2006). In addition to finding evidence for straital-based prediction error, den Ouden et al. (2009) found evidence for predictive coding in the primary visual cortex. Specifically, the visual cortex responded most when an un-

expected visual stimulus was presented but also when an expected visual stimulus was not presented. This latter response is not predicted by a feature-based theory of cortical responses (e.g., Grill-Spector et al., 2006) but is consistent with Predictive Coding models of cortical response (e.g., Friston, 2005). Parallel results are reported by Egner et al. (2010); this study found that activity in the ventral visual stream responds differently based on the relatively predictability of categories of complex visual stimuli. The probability that a face or a house stimulus would be presented was manipulated by a cue.[4] According to models of Predictive Coding, activity in the ventral visual cortex will not respond simply based upon the features of the given stimuli but based on whether that type of stimulus is predicted. Figure 6.2 provides an illustration of the cortical responses predicted from a Predictive Coding model and standard, feature-based models of cortical response. Activity in a face-selective area of the visual cortex (fusiform face area, FFA) was found to be modulated by the probability of receiving a stimulus from a given category (see Figure 6.3). Taken together, these results support a view that responses in perceptual cortices are predictive in character and modulated by the relative predictability of sensory input. While future research is necessary to establish the extent to which these effects generalize to other statistical-based incidental learning tasks such as canonical statistical learning paradigms, evidence of predictive coding during incidental learning suggests that statistical learning could be supported by an active, prediction-based learning process.

In sum, considering the active-passive dichotomy presented in the beginning of this section, statistical learning and other forms of incidental learning

[4]Participants were asked to perform a cover task of categorizing occasional inverted stimuli. This task is arguably an irrelevant behavioral task as the target inverted stimuli were not manipulated by the cue and was orthogonal to the perceptual expectation manipulation. Thus, Egner et al. (2010) is also passive as defined in the beginning of the section.

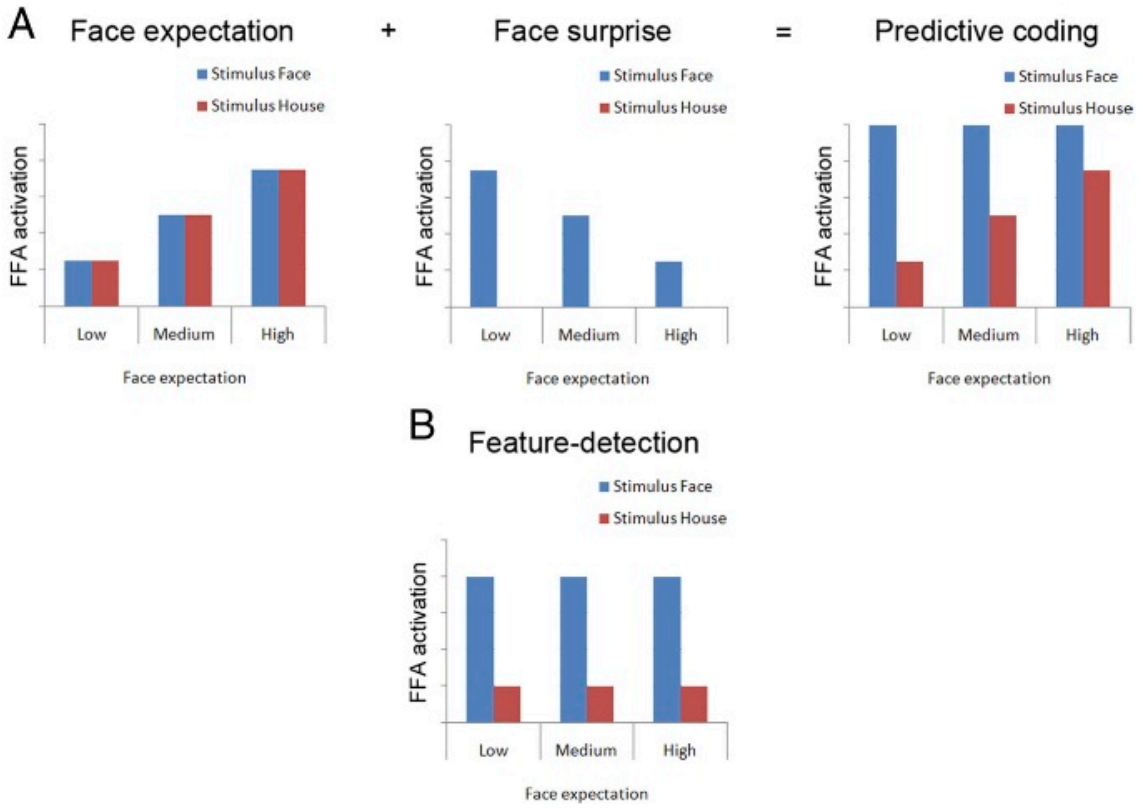


Figure 6.2: From Egner et al. (2010), graphs depict the fusiform face area (FFA) population response patterns predicted by the Predictive Coding and feature detection theories of cortical function. A, Predictive coding argues that FFA population responses reflect the sum (right) of activity generated by representation units (face expectation, left) and error units (face surprise, middle). B, Feature detection views suppose that the FFA population response is driven by stimulus features, with face stimuli eliciting stronger responses than house stimuli.

are certainly passive in so far as they do not involve task-guided behavior or feedback. However, this section argues that the cognitive processes underlying incidental learning, and statistical learning specifically, can be conceived to be active in that they involve prediction of successive stimuli and prediction error rather than the passive absorption of statistical information. This section both presented behavioral results that are inconsistent with such a passive view of the mechanisms supporting statistical learning and evidence for both predic-

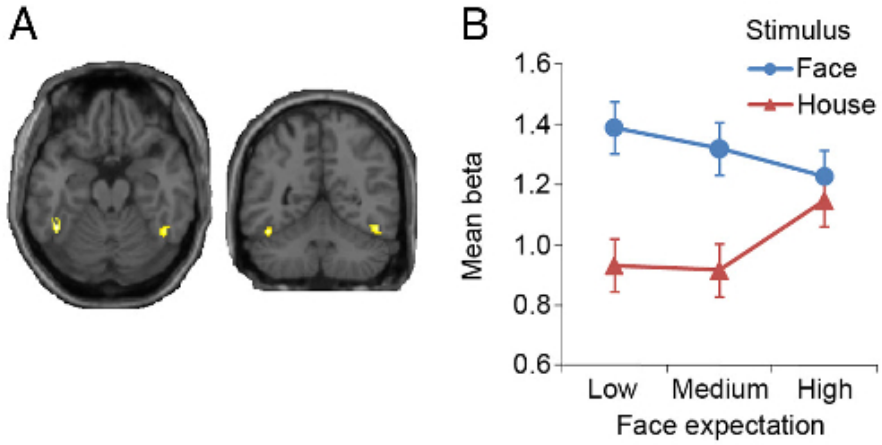


Figure 6.3: From Egner et al. (2010), A, FFA localizer group results and B, mean group beta estimations for the ROI defined by the FFA localization displayed in A.

tion error and predictive coding in incidental learning tasks driven by relative predictability of sensory input.

6.2.2 Information Gathering and Statistical Regularities

Even if incidental learning, such as statistical learning, and more active forms of learning, such as reinforcement learning, have some overlapping learning mechanisms, it is an important area for future research to determine how incidental learning and more behaviorally-active learning interact to support developmental change. There is some indication that experience with statistical regularities might affect behaviorally active forms of learning and specifically with active information gathering or exploration. Goldstein et al. (2010) proposes an interaction of statistical significance, to be picked up incidentally, and behavioral-significant, such as reinforcement through social interaction, in learning from structure. Thus, an interaction of incidental and active learning

could bias the process of learning structure in general and specifically, language acquisition. Building on this example, this subsection explains the role of exploration in developmental change according to the Ecological perspective and reviews some evidence that experience with regularities can affect such information gathering activities.

Exploration is an important tool at the disposal of the observer and supports changes in perception as a result of experience. The layman's definition of exploration entails an overt, conscious searching of the environment. By contrast, Ecological approaches emphasize a connection between perception and exploration. Exploration does not occur after perception but in the service of perception. Thus, exploration can be viewed as a form of directed information gathering (E. J. Gibson, 1969).

There are many possible avenues for exploration or information gathering. For example, in her first year of post-natal life, an observer could employ manual and oral exploration, reaching, sucking, crawling, attentive listening, etc. One key method for sensory exploration, available to the newborn and important throughout the lifespan, is eye gaze. In fact, within the first few months of post-natal life, infants make millions of eye movements (Haith et al., 1988). Recent research has linked changes in eye movements to changes in visual object perception (Amso & Johnson, 2006; S. Johnson et al., 2004). This research emphasizes that exploration of sensory input changes the resulting perception of that sensory input and such information gather abilities increase throughout development and with experience (E. J. Gibson, 1969; E. J. Gibson & Pick, 2000).

There is some evidence that information gathering behaviors can be affected by the presence of predictive regularities in sensory input. Haith et al. (1988)

found that experience with statistical regularities supported changes in exploratory eye movements in 3.5-month-old infants. As reviewed in Section 1.3.1, infants were presented with either a predictable sequence of visual stimuli or an unpredictable sequence. Even though infants had no control over the visual stimuli in both conditions, infants viewing the predictable sequence exhibited more anticipatory eye movements (eye movements produced before the presentation of the visual stimulus) and had faster facilitated eye movements (eye movements elicited after the presentation of the visual stimulus). This study specifically finds a relationship between experience with statistical regularity and infants' exploration.

This dissertation also presents empirical evidence that experience with environmental regularities can shape exploration strategies. Chapter 5 examined eye movements during experience with structured sensory input (Experimental Condition) compared to sensory input without any supportive structure (Control Condition). Eye movements to the Target Scenes to examined in relation to exposure condition but also whether a participant changes their post-test perception of the Target Object. We found that the strongest predictor of eye movement patterns is exposure condition: participants in the Experimental Condition had markedly different patterns of fixation compared to the Control Condition regardless of whether they changed their perception. We do also find evidence of modest differences in eye movements for participants who change their perception in the Experimental Condition and specifically, differences in visual exploration of the Target Scene. These results build on research in infants showing that individual differences in visual exploration predict individual differences in visual object perception and suggest that experience with statistical regularities in sensory input might facilitate changes in exploration strategies and in

turn support changes in perception.

A relationship between incidental experience with statistical regularities and visual exploration might be a powerful mechanism to support developmental change. Haith et al. (1988) emphasize that such a relationship will facilitate the development of a number of important behaviors including motor development, which depends on visual monitoring, such as reaching and catching but would likely also support behavioral-interaction with visual objects in general. A similar argument is made by Fernald and colleagues. A number of studies have found that childrens' visual exploration can be affected by predictive relationships in language (presumably learned through incidental language experience) such as the first syllable of a word or the marker of grammatical gender which is spoken before a noun. Fernald and colleagues have argued that faster visual exploration based on predictive relationships in language is particularly important to the development of language comprehension given that language is a fast-paced temporally unfolding and transient signal (also reviewed in Section 1.3.1, see Fernald et al., 2001; Lew-Williams & Fernald, 2007).

As asserted by Piaget (1971) "anticipatory function... is to be found over and over again at every level of the cognitive mechanisms and at the very heart of the most elementary habits, even of perception" (p. 19). The current section argues that incidental learning from statistical regularities can be considered active in the sense that it likely involves both predictions and prediction error. Despite any overlap between the mechanisms supporting behaviorally-active and behaviorally-passive learning, the interaction between these types of learning may uncover powerful mechanisms for developmental change. As an example, the author outlines evidence that experience with statistical regulari-

ties can support changes in eye movements. It is important to note that given that this mechanism depends on the mechanisms of visual selective attention, there will be cognitive limits on the effectiveness of this mechanism early in the first post-natal year, see Section 5.3. However, the possible relationship between incidental learning and information gathering strategies remains an important area for future study.

6.2.3 Sensory Predictability Can Affect Perception

This dissertation provides some initial evidence for the claim that learning and memory systems can shape perception as a result of experience with regularity in sensory input. Briefly, Chapter 5 found that structure present in sensory input can drive changes in object perception through the involvement of both the medial temporal lobe, likely binding across variable experiences to extract invariant representations of the novel Target Object, and fronto-striatal circuity which may drive changes in eye movements seen during the Experimental condition. Thus, structured sensory input can affect perception through activity in learning and memory systems. Moreover, this mechanism is available in infancy and could potentially drive changes in object perception early in the first post-natal year (see Section 5.3)

While the results in Chapter 5 present one mechanism by which experience with regularity in sensory input can affect changes in object perception through the involvement of learning and memory systems, this dissertation suggests that this specific mechanism by which learning and memory systems can affect perception is simply one avenue by which learning and memory systems can

affect perception. While much future empirical work is needed in this vein, this section builds on the proposal that statistical learning is the result of prediction-based processes and suggests a second avenue by which experience with statistical regularities can affect perception.

Differences in predictability of sensory input have been shown to modulate cortical responses in relevant perceptual regions and cortical connectivity between regions. As reviewed in the previous section, recent studies have found that when predictability of given stimuli is modulated (e.g., the associative strength of a visual stimulus following an auditory cue or the relative predictability of a certain category of visual stimulus) that perceptual cortices respond in accord with this predictability (den Ouden et al., 2009; Egner et al., 2010). den Ouden et al. (2009) also found that the relative predictability of a visual stimulus following an auditory stimulus affects the cortical connectivity between primary auditory and visual regions. Specifically, dynamic causal modeling techniques revealed that the strength of an A1 $\rightarrow$ V1 connection changed as a function of the associative strength predicted by the R-W model. Together these findings suggest that functions of perceptual systems can be affected by the relative predictability of sensory input.

Consistent with the proposal put forward in this dissertation, recent evidence has suggested that these changes in activity of perceptual systems are the result of influences from learning and memory systems. den Ouden et al. (2009) found some indirect evidence that the cortical connectivity between primary auditory and visual regions was mediated by the striatum: A dynamic causal modelling found that the superior model included a single mediating connection between A1 and V1; a parallel pattern of activation in the striatum

suggested that the striatum was the mediator between A1 and V1. If the striatum was indeed the mediating connection, its function would be to adjust cortical connectivity between these regions based upon prediction errors and affect predictive coding in V1. Extending these results, den Ouden, Daunizeau, Roiser, Friston, and Stephan (2010) employed a cross-modal associative learning task, similar to the one employed in den Ouden et al. (2009). However, in this follow-up study, participants were asked to discriminate amongst two types of visual stimuli (faces and pictures of places). These two types of stimuli were preceded by different auditory cues. Specific auditory cues predicted specific types of visual stimuli with different levels of strength. The relationship between cues and targets varies dynamically across the experiment requiring participants to continuously update the predictive relationship between acoustic and visual stimuli. Trial-by-trial changes in associating strength were modeled using a Bayesian learning model. As with den Ouden et al. (2009), striatal activity was found to approximate the prediction errors for both visual stimuli and two distinct regions of the visual cortex corresponding to the fusiform face area and the parahippocampal place area also exhibited predictive coding. Critically, den Ouden et al. (2010) found that prediction errors in the striatum modulated connectivity between these stimulus-specific visual areas and the motor regions supporting the discriminative response (premotor cortex). This result extends the findings of den Ouden et al. (2009) in two important ways: First, this study provides direct evidence that striatal prediction error is used to tune functional connectivity in cortical networks, and specifically, connectivity with perceptual cortices; second, these results provide some evidence that these trial-by-trial changes in predictability and effects of predictive coding have some behavioral efficacy as predictability was associated with response speed and accu-

racy. Thus, the results from den Ouden et al. (2009) and den Ouden et al. (2010) suggest a second avenue by which learning and memory systems can translate experience with statistical regularities into changes in perception. Specifically, prediction errors produced in the striatum can be used to modulate cortical connectivity with stimulus-specific regions of neocortex. However, future research is needed to determine whether such modulations of cortical connectivity and predictive coding can produce lasting changes in perception before these results can be more directly applied to the lasting experience-based changes in perception that this dissertation aims to explain.

This dissertation proposes that experience can shape perception is through the response of learning and memory systems to statistical regularities. This proposed framework is potentially powerful given the ubiquity of statistical information in sensory input. Chapter 5 presents a specific mechanism by which variable visual experience with objects can support changes in object perception through activity in multiple learning and memory systems (MTL and BG). This section also outlines a second mechanism by which prediction error produced in the striatum can affect activity in perceptual cortices and changes in stimulus detection.

6.2.4 Learning and Memory Systems Can Affect Perception: Conclusions

Perceptual processing has typically been conceived as a passive detector of information in the environment. The studies reviewed in this section suggest that perceptual cortices, as well as learning and memory systems, can engage in pre-

dictive coding and produce prediction errors in response to experience with statistical regularities. The involvement of active predictive processes in response to statistical information presents a more active view of perception and one that is shaped by experience, in accord with the Ecological views presented in Chapter 1. This section discusses empirical evidence both from this dissertation (Chapter 5) and from other recent findings (e.g. den Ouden et al., 2010) that learning and memory systems are involved in translating experience with statistical regularity into adaptive changes in perceptual processes. While currently the evidence for the direct impact of learning and memory processes on changes in perception is sparse and there are many areas remaining for future study in order to support this argumentation, these results support the broader framework presented in this dissertation that learning and memory system and perceptual systems are mutually influential to support adaptive experience-based change in perceptual processing throughout the lifespan.

6.3 Implications for Developmental Psychology

The field of developmental psychology studies systematic psychological changes that occur across the lifespan. The current empirical work examines learning from statistical regularities in sensory input. Statistical learning has been implicated in the development of a number of perceptual-cognitive tasks and, most notably, language acquisition (e.g., Saffran & Thiessen, 2007). Despite a number of studies demonstrating statistical learning using many types of sensory input and some consensus that statistical learning is involved in developmental change, very few studies have helped to elucidate the nature of the learning mechanism(s) supporting statistical learning and how these mech-

anisms operate given the complex and variable sensory experience that drives language development (Romberg & Saffran, 2010). This dissertation contributes a novel view to both of these important open questions. Indeed, a greater understanding of both the mechanisms of statistical learning and how these mechanisms can operate given real-world sensory input is the major contribution of this dissertation to the field of developmental psychology. See Section 6.1 for an in-depth discussion on this topic.

The empirical evidence presented in this dissertation is broadly consistent with a number of contemporary approaches to cognitive development. While the goal of the current work is not to take a stance on the origins of human knowledge, the specific emphasis on experience with statistical information found in sensory input in developmental change is most consistent with empirical approaches to development and de-emphasizes the need for innate knowledge in adaptive developmental change, as proposed by nativist theories. Currently there are a number of developmental theories or frameworks that emphasize the role of experience-driven changes in perception and cognition, including Connectionism (e.g. Elman et al., 1996; Thelen & Bates, 2003), Bayesian inference (e.g. Gopnik & Tenenbaum, 2007; Tenenbaum et al., 2011), Neuroconstructivism (Mareschal et al., 2007) and Neo-Constructivism (S. Johnson, 2010). These frameworks have different foci and the current work has overlapping emphasis with all of these approaches. For example, Neuroconstructivism emphasizes the interrelationship between neural and cognitive development, while Connectionism emphasizes the need for prediction error in driving changes in representations.

The major division within the contemporary theoretical approaches, indi-

cated above, is between constructivist approaches (e.g., Neo-constructivism, Neuroconstructvism and Bayesian inference) and associationist approaches to learning and development (e.g. Connectionism). While this dissertation remains generally neutral on this theoretical division, current results do weigh on this debate within the area of visual statistical learning. Constructivism emphasizes changes in knowledge as a result of experience. For example, Neuroconstructivism emphasizes that development results in changes in the complexity of representation and Bayesian inference emphasizes a hierarchical integration of knowledge. Associationism emphasizes that developmental change can occur through changes in connections as a result of experience. Associationists criticize Constructivist approaches for building in abstract, complex representations, and conversely, Constructivist approaches criticize Associationism for failing to capture the complexity of human cognition including abstract thinking and ability to generalize knowledge. Some recent work has sought to determine whether patterns of statistical learning is more consistent with Associationist or Constructivist (specifically, Bayesian) approaches. Orbán, Fiser, Aslin, and Lengyel (2008) presented evidence that the ability of human observers to chunk visual elements into single multi-element entities is captured by Bayesian models but not associationist models. These behavioral findings are reminiscent of the results presented in Chapter 5 where the consistent association of two parts of a single object in a scene did not support a unified perception (in the Control Condition), but variable experience with the Target Object in additional scenes did support a change to a completed percept (in the Experimental Condition). This result fits well with Bayesian and, by extension, Constructivist views of experience-based perceptual-cognitive change.

The goal of the current work is not to refute or support specific theoretical

views; however, this work generally supports an empirical view that development can result from experience and specifically, experience with sensory input endowed with statistical regularities. This view is broadly consistent with a number of contemporary approaches to development. Instead, the implications of this dissertation to developmental psychology is to provide some insight into the mechanisms by which experience with statistical information in sensory input can result in adaptive changes in perception.

6.4 Implications for Cognitive Science and Cognitive Psychology

As summarized in the beginning on this chapter, a major focus of contemporary cognitive science, and specifically cognitive psychology, is elucidating the nature of cognitive processes and the representations they act upon. An important component of this pursuit is the debate about which, if any, mental processes are modular or informationally-encapsulated and which processes are informationally-continuous (e.g., Fodor, 1983; Spivey, 2007). This section will discuss the implications the results presented in this dissertation have for this debate.

The modularity debate includes the nature of perceptual processes: Some researchers defend a modular view of perception, while others assert that perception is continuous with cognition. This section will focus on the modularity debate as it is relevant for visual perception, but the theoretical views are relevant to questions concerning all perceptual modalities. One can frame the modularity debate as answers to the question "Why do we perceive the world

the way that we do?" and contrast answers. Considering visual perception,

> the question of why we see things the way we do in large measure still eludes us: Is it only because of the particular stimulation we receive at our eyes, together with our hard-wired visual system? or is it also because those are the things we expect to see or are prepared to assimilate in our mind? (Pylyshyn, 1999, p. 341)

If the former answer, that we see things the way we do because of the pattern of sensory input we receive, is consistent with a modular view of visual perception: visual perception is affected only by the visual sensory input provided and the static response of the system to that input (see Pylyshyn, 1999). The latter answer is consistent with a continuous view of perception and cognition. Visual perception is permeable to cognitive functions such as expectations. Following from Bruner (1957), a common line of evidence for this view is the permeability of perception based on motivation and experience (see a nice summary of this field by Pylyshyn, 1999). Recent research in this vein by Balcetis and colleagues has found that an observer's political partisanship affects an observer's perception of a candidate's skin-tone (Caruso, Mead, & Balcetis, 2009), and that desired objects are seen as closer than undesirable objects in both visual estimates of distance and in their motor responses (bean bag tosses) to the object (Balcetis & Dunning, 2010). Similarly, evidence for top-down influences in visual perception also support a continuous view of visual perception. A classic example of this is perception of Mooney faces. Mooney faces are impoverished images of faces that are extremely difficult to recognize (Mooney, 1957). However, primes of less impoverished visual images of the faces supports rapid and robust perception of the impoverished Mooney faces. Recent research with

Mooney faces has also found evidence that prior experience with a face (e.g., a famous face such as Albert Einstein) supports differences in the perception of impoverished Mooney faces of that individual (Jemel, Pisani, Calabria, Crommelinck, & Bruyer, 2003).

This dissertation argues for an informationally-continuous view of perception and cognition. The author argues for an informational interrelationship between perceptual and learning and memory systems. Building on the research reported in the previous paragraph, the study in Chapter 5 finds evidence that activity of learning and memory systems during variable yet regular experience with a novel object can change the perception of that object, without any changes in the sensory input (i.e., the Target Scene is identical both across experimental conditions and throughout the experiment but experience modulates changes in perception of the Target Object in this scene). The finding that perception can change without changes in sensory input and that these changes are dependent upon the type of visual experience refutes a modular view of visual perception as outlined above and in Pylyshyn (1999).

In sum, an important debate in Cognitive Science is whether perceptual processes are informationally-encapsulated from other aspects of cognition. The author argues that perceptual processes are not modular or encapsulated but rather are informationally-continuous with other aspects of cognition and specifically learning and memory systems. Building on evidence that perception changes with motivation and top-down modulation, findings in Chapter 5 provide evidence that representations created in part by the activity learning and memory systems can change perception.

6.5 Implications for Cognitive Neuroscience

The field of Cognitive Neuroscience investigates the relationship between mental or cognitive and neural processes. This dissertation has argued for an interrelationship between perceptual processes and learning and memory processes during experience with novel statistical regularities in sensory input. In the field of Cognitive Neuroscience, specific neural regions are associated with these cognitive processes; occipital cortices is believed to support visual processes and the medial temporal lobe is believed to support learning and memory processes. Recent research has challenged the exclusive association between activity in the medial temporal lobe and learning and memory and proposed that some regions of the medial temporal lobe might support perceptual functions. The current section discusses how the framework presented in this document may influence this ongoing debate in the field of Cognitive Neuroscience.

While compelling and convergent evidence exists that the medial temporal lobe is involve in learning and memory functions, recent research has suggested that regions in the medial temporal lobe are also important for perception. Some of the early evidence for continuity between visual perception and learning and memory was provided by studies of monkeys with perirhinal cortex lesions. Specifically, these monkeys exhibited a deficit in an oddity-discrimination task (Buckley et al., 2001). In this task, a number of visual stimuli are presented simultaneously with one "odd" stimulus; see Figure 6.4. Monkeys with perirhinal cortex lesions showed deficits in oddity-discrimination only when the stimuli were complex, as illustrated in this figure (see Baxter, 2009; Suzuki, 2009, for other compelling results). Accordingly, some researchers have suggested that rather than a clear distinction between the perceptual functions of the ventral

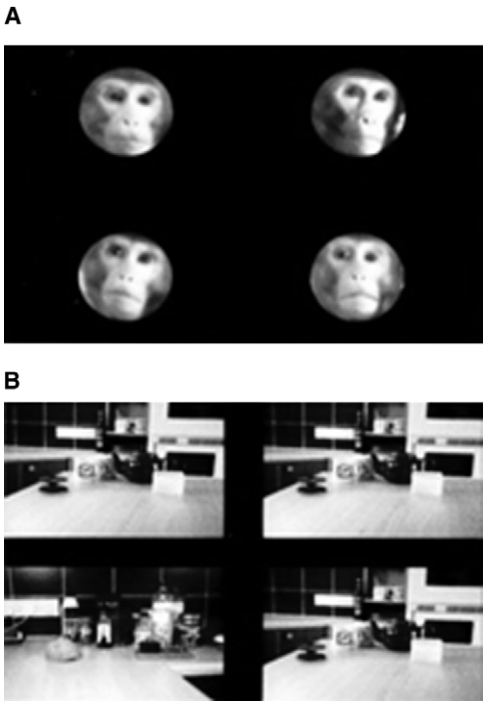


Figure 6.4: An illustration of the oddity-discrimination task employed by Buckley et al. (2001). In this task, an animal must identify the odd stimulus in an array to receive a reward. Panel A presents illustration of the monkey face oddity task. The face pictured in the top right corner is the odd face. Panel B presents a scene oddity task. The scene in the lower left corner is the odd scene.

visual cortex (e.g. visual object processing, Gauthier et al., 1999; Kanwisher et al., 1997; Logothetis & Sheinberg, 1996) and the learning and memory functions of the medial temporal lobe (e.g., Eichenbaum & Cohen, 2001), there exists a continuum between visual perception and learning and memory with some parts of the medial temporal lobe contributing to some aspects of perception in addition to learning and memory functions.

However, one of the outstanding issues in this debate is how to determine if higher-level visual tasks are purely perceptual or involve some learning and memory. Returning to the oddity-discrimination task developed by Buckley and colleagues and summarized in the previous paragraph, this task is believed

to be a result of perceptual processes and not learning and memory processes because all stimuli are presented on the screen simultaneously. However, these tasks could also be considered to involve some aspects of associative learning and memory processes. Specifically, as argued by Suzuki (2009), "[t]o successfully identify the odd stimulus in this task, animals needed to associate the different views of the same face as one"(p. 659) as illustrated in Panel A of Figure 6.4. Thus, the field is unclear about what comprises a perceptual task and when a task conflates learning and perception.

This dissertation argues against a clear distinction between the process of perception and the processes of learning and memory. Again, the results presented in Chapter 5 suggest that activity in the medial temporal lobe is important for adaptive changes in visual object perception as a result of experience. The author does not argue that there is no learning component in the experiment employed in Chapter 5 and correspondingly that the activity of the hippocampus associated with perception is evidence for medial temporal lobe involved in purely perceptual tasks. Adopting a different approach, the author suggests that instead of searching for a clean division between what tasks are perceptual with no aspect of learning and memory and *vice versa*, it might be fruitful to examine to what extent the medial temporal lobe, already associated with learning and memory, are also involved in perceptual tasks and to what extent cortices already associated with perceptual processes are involved in learning and memory function. It has already been observed that perceptual tasks that involve a high degree of ambiguity and featural overlap involve the medial temporal lobe (Suzuki, 2009). It is important to question why such tasks benefit from the involvement of medial temporal lobe structures, which other tasks with a perceptual component also recruit these regions, and why some perceptual tasks do

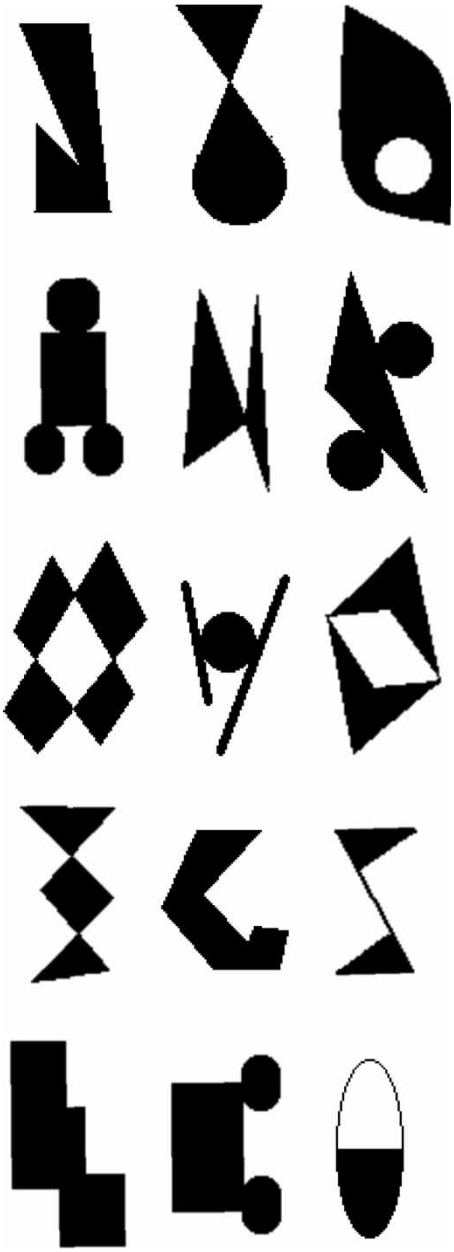
not benefit from the involvement of learning and memory systems. In sum, the author suggests that instead of pursuing a distinction between perception and learning and memory contributions in high-level visual tasks, it might be fruitful to take a different approach as presented in this dissertation: to investigate how these systems interact during vision tasks to support robust and adaptive internal representations of the external world.

To conclude, there is an ongoing debate as to whether regions of the medial temporal lobe are involved exclusively in functions of learning and memory or whether they are also involved in high-level visual perception. Evidence presented in this dissertation suggests that learning and memory systems may aid in the shaping of visual perception and specifically object perception. However, the author suggests that while the question of how to cleanly dissociate the learning and memory from high-level vision is an interesting philosophical question, investigating the interaction of these systems might lead to a deeper understanding how the brain creates a robust, adaptive and rich internal representation of the external world.

APPENDIX A

FIFTEEN SHAPES USED IN ALL EXPERIMENTS IN CHAPTER 3

GROUPED INTO ARBITRARY TRIPLETS



APPENDIX B

MONOSYLLABIC NONWORDS USED AS AUDITORY STIMULI IN CHAPTER 3

The 225-ms monosyllabic nonwords used in Experiments 1 and 3A

bu, cha, da, el, feng, jic, meep, rau, roo, rud, sa, ser, ta, wif, zet

The 450-ms monosyllabic nonwords used in Experiments 2 and 3B

bu, cha, dak, eeg, feng, jeen, jic, meep, pel, rauk, rous, rud, sa, ser, wif

APPENDIX C

TESTING FOR VISUAL MASKING AT FAST RATES OF PRESENTATION

Given the comparably reduced temporal resolution of the visual modality, it is important to test whether the reduction in visual ISL at faster rates of presentation is due to a reduction of discriminability in the visual modality and not due to a reduction in learning due to temporal presentation more generally. In order to test this possibility, we compared the accuracy at making simple judgments at the two rates of presentation employed in the current experiment. To confirm that there are no differences in simple discrimination across modalities and rates of presentation, the current experiment has participants perform a same-different judgment using both visual and auditory stimuli at both rates of presentation.

C.1 Method

Participants.

Twenty participants were recruited from Psychology classes at Cornell University earning extra credit. All participants reported normal or corrected-to-normal vision, no serious auditory deficits, or neurological problems.

Materials.

The same 15 visual shapes were used as in the experiment reported in the main text, as were the two sets of non-words employed at each presentation speed. For each participant, for each task, a different subset of 5 visual or 5 auditory

stimuli was used in the control task. Over all participants, each stimulus within each set of stimuli was paired with every other stimulus from the same set.

Auditory and visual stimuli were presented at the same duration as they were presented during the main experiment (750ms SOA or 375ms SOA), with a 1000ms blank screen before and after the presentation of the two successive stimuli.

Procedure.

Trials were blocked according to presentation speed and modality of presentation; blocks were presented in random order. Within each block, each shape or non-word of the subset was paired with every other as well as with itself to form 60 trials and presented in random order. Participants were simply asked to report whether the two stimuli they saw were "the same" or "different" by pressing keys 1 and 4, respectively, on a 4 button box.

C.2 Results and Discussion

In all groups, participants responded with high accuracy: visual, slow presentation: mean = 99.3%; visual, fast presentation: mean = 99.7%; auditory, slow presentation: mean = 98.4%, auditory, fast presentation: mean = 99.2%. Planned paired t-tests were conducted to directly compare performance a) within modality for each presentation speed and b) within each presentation speed for each modality. None of the above comparisons yielded significant differences between conditions: auditory (slow vs. fast presentation): $t(20)$ = -1.56, $p > 0.1$;

visual (slow vs. fast presentation): $t(20)$ = -0.55, $p > 0.5$, slow presentation (auditory vs. visual modality): $t(20)$ = 0.641, $p > 0.5$, fast presentation (auditory vs. visual stimuli): $t(20)$ = 1.101, $p > 0.25$. A univariate ANOVA confirmed that there is no main effect of either presentation speed nor modality and no interaction between factors.

If masking did indeed obfuscate the perception of visual stimuli at the faster rates of presentation, simple identity judgments should decline at faster presentation rates. The current results confirm that there is no difference in simple discrimination of the stimuli employed in the current experiment either between modalities (auditory vs. visual) and/or at different rates of presentation (fast vs. slow). Thus, the supplementary results support our main thesis that temporal perceptual grouping factors interact with learning beyond simple differences in stimuli discrimination across timing conditions.

APPENDIX D

STAYING GROUNDED: IS STATISTICAL LEARNING CONSTRAINED TO LESS ABSTRACT PATTERNS?

The content of this chapter is under review at *Cognitive Science*; see Emberson and Rubinstein (under review)

This paper examines statistical learning in the presence of regularities at multiple levels of abstraction. We presented participants with pictures of objects where picture order was predicted by both object identity and object category. In Experiment 1, we establish that participants learn based on object-specific regularities when regularities exist at both levels of abstraction. In Experiment 2, we investigate whether participants engage in additional category-level learning and find evidence that learning is constrained to less abstract, object-specific patterns. In both experiments, we find equal learning between participants who viewed typical and atypical exemplars suggesting, again, that participants do not gain category-level knowledge. Overall, our results indicate participants preferentially learn based on object-specific regularities over more abstract, category-level regularities in a statistical learning task. These findings suggest a novel constraint on learning from environmental statistics, and a possible direction of learning across multiple levels of abstraction in a complex environment.

D.1 Introduction

Throughout our lifetime, experience shapes our mental model of the world. Learning from patterns, regularities or statistics found in the environment constitutes one way to shape cognition as a result of experience. Despite the clear importance of learning from patterns in the environment, there are many outstanding questions about the nature of the mechanisms that support learning from real-world experience. A central problem in this literature is how learning mechanisms operate given the richness of the information we get from the world. Are learning mechanisms a priori constrained to learn particular patterns? This is especially difficult given the multiplicity of ways in which an object can be represented in a cognitive system. This paper examines whether learning mechanisms are biased towards patterns of low levels of representational abstraction or higher-level, more abstract patterns.

We focus on a type of environmental learning called *statistical learning* where participants passively learn from stimuli embedded with probabilistic information. Previous research has supported the view that these experiential learning mechanisms are largely unconstrained: statistical learning has been demonstrated in multiple sensory modalities (Conway & Christiansen, 2005) and across a wide range of perceptual input. For example, in the visual modality, learning can occur from sequences of gestures (Baldwin et al., 2008) or simple shapes (Fiser & Aslin, 2001). While the majority of these studies have focused on learning probabilistic relations of individual objects, there is evidence that learning can occur at higher levels of informational abstraction including based on categories of nonsense words (Saffran, 2002) or familiar semantic categories (Brady & Oliva, 2008). Overall, these studies support the view that environ-

mental learning is unconstrained. That is, if there is any reliable probabilistic information in the environment, humans can learn from it regardless of level of abstraction or perceptual properties.

However, these demonstrations of unconstrained learning arise from paradigms where environmental regularities exist only at a single perceptual and/or informational dimension. To illustrate, Brady and Oliva (2008) use a paradigm where the categories of scenes are predictive of picture order but individual scenes are not (e.g. beaches predict kitchens as categories of scenes but beach_1 does not predict kitchen_1). In this paradigm, it is not possible to learn based on individual scenes because only category-level regularities are present. Thus, these studies provide an existence proof of an unconstrained learning mechanism but they arise under specific, restricted conditions.

In actuality, the learners environment is characterized by regularities at many levels of abstraction observed simultaneously and often redundantly. For example, the predictive relationship between dogs and leashes exists based on abstract categories as well as in the actual objects or exemplars seen in the world (e.g. dogs have their specific leashes). Reflecting this key aspect of everyday experience, the current paper examines learning when participants are exposed to environmental regularities at multiple levels of abstraction. Do participants learn from the *multiple levels* of predictive dependencies simultaneously or are they biased to learn at a certain level of abstraction?

To address these questions, we devised a novel statistical learning task where predictive regularities are learnable and redundant across two levels of abstraction. Specifically, participants were presented with sequences of previously unseen exemplars from known basic-level categories. Both the categories

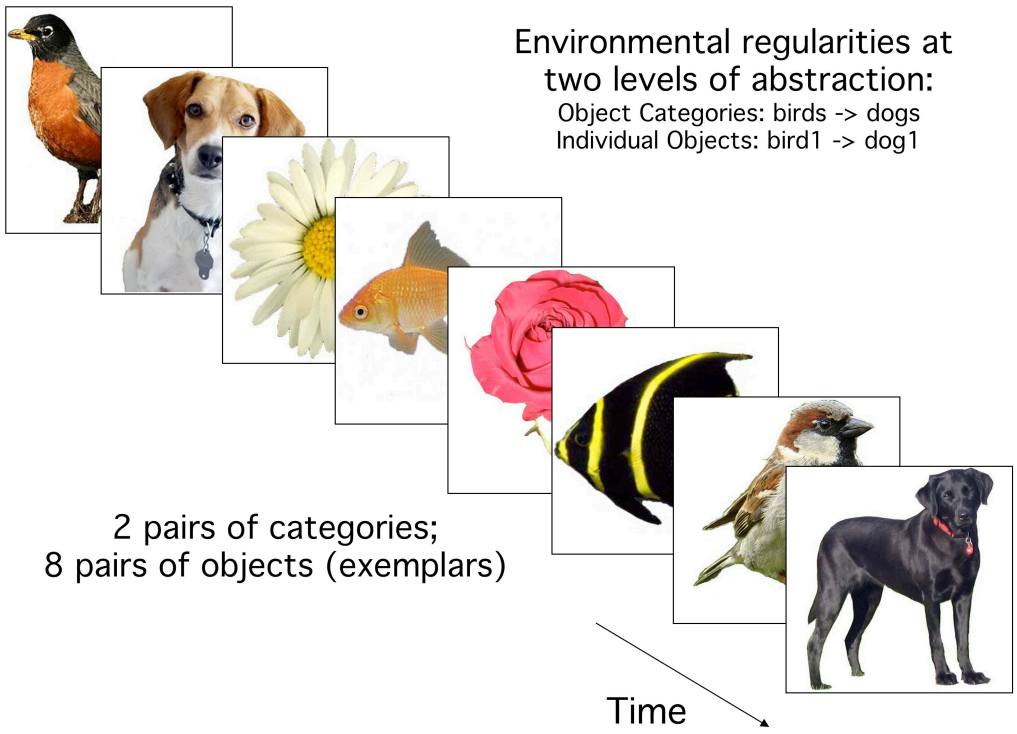


Figure D.1: A sample familiarization stream. The same familiarization was given to participants in both experiments. Pictures were organized into pairs of categories (e.g. birds → dogs) as well as specific objects within these categories (e.g. robin → beagle). Thus, predictive regularities were redundant across multiple levels of abstraction resulting in two pairs of categories and eight pairs of objects or exemplars of these categories. In the sample stream, birds predict dogs and flowers predict fish.

(e.g. dogs-fish, flowers-birds) and the individual exemplars of these categories (e.g. dog_1-$fish_1$, dog_2-$fish_2$) were predictive of picture order (see Figure D.1).

This experimental design provides ample opportunity for learning at the abstract, categorical level. First, previous research has established that the categories used in the current experiment are initially processed at the basic-level, the level of categorical regularities in the current task, rather than the subordinate level (e.g. dog as opposed to the subordinate level of beagle; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). Second, we employed the same stimu-

lus timing (short durations and long inter-stimulus-intervals) as employed by Brady and Oliva (2008) which likely taps into the fast, gist-based recognition of the pictures. Finally, the stream has fewer pairs of categories than objects (see Figure D.1). Thus, category level learning is, in some sense, easier than object-specific learning. Moreover, using the same methodology as Brady and Oliva (2008) and described above, pilot testing confirmed that participants can learn from categorical regularities when they are the sole predictor of picture order: mean = 62.6%, $t(13) = 2.80$, $p < 0.05$. These results confirm that if object-based regularities are not present, category-level learning is possible using the current stimuli.

In addition, we manipulated object typicality, providing another means to examine learning at the category level. Atypical objects are processed differently from typical objects (Dale, Kehoe, & Spivey, 2007) and tend to be more quickly processed below the basic-level categories (e.g. penguin as opposed to bird; Jolicoeur, Gluck, & Kosslyn, 1984). Thus, we expect the participants familiarized with atypical exemplars to have weaker category-level learning but equivalent learning at the object-specific level. Figure D.2 presents the atypical exemplars, while all other figures present typical exemplars.

D.2 Experiment 1: Testing for Object-Level Learning

We first examined whether participants continue to be sensitive to object-specific regularities when more abstract regularities are present. Employing a well-established testing procedure (e.g., Brady & Oliva, 2008; Fiser & Aslin, 2001), participants were asked to distinguish pairs of pictures from familiariza-

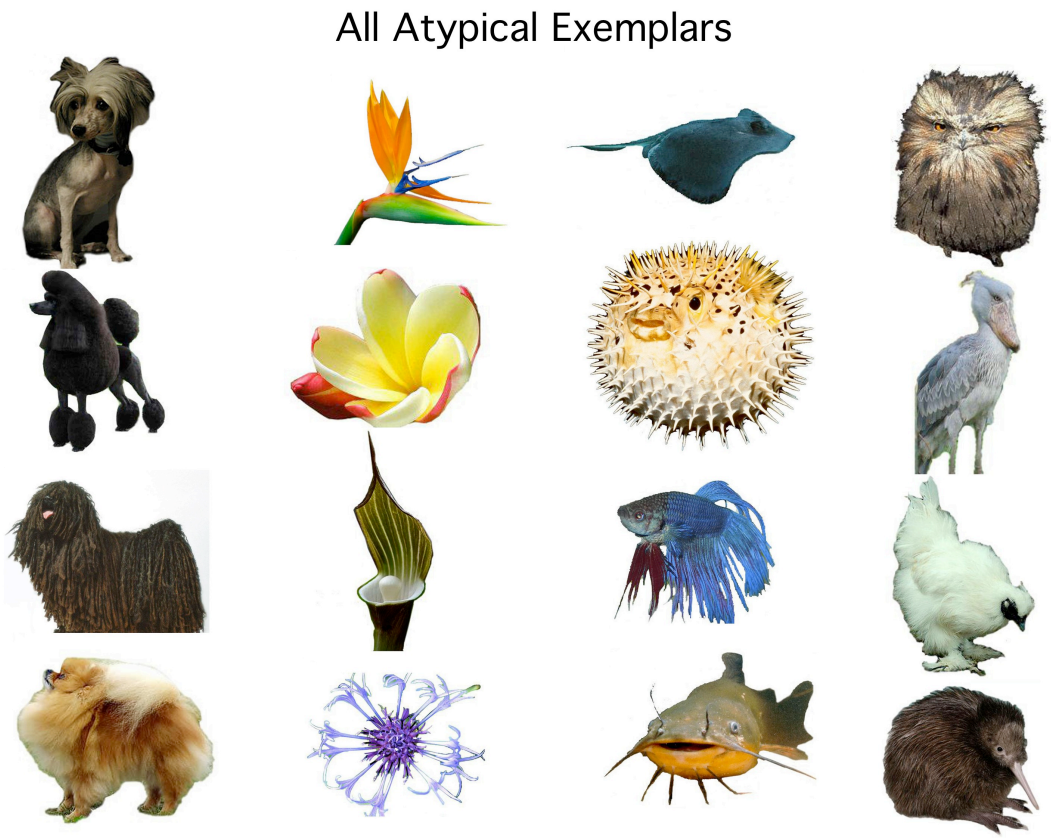


Figure D.2: All atypical exemplars used in the current paper, organized by category (from left: dog, flower, fish, bird). Half of participants received familiarization with atypical exemplars, pictured here, and half were familiarized with typical exemplars depicted in Figures D.1 and D.3. The test was conducted with the same pictures as familiarization. We hypothesize that any category-level learning will be modulated by typicality of exemplars with atypical exemplars resulting in weaker category-level learning.

tion (e.g. $bird_1$-dog_1) from a foil pair created from the same pool of pictures but which violated contingency patterns of the familiarization stream. To isolate knowledge at the object-specific level, the foils were designed to violate object-based regularities while maintaining categorical regularities (e.g. $bird_1$-dog_2, see top panel of Figure D.3). Since participants require object-specific knowledge of the familiarization stream in order to distinguish the foils from the pairs, the ability to distinguish pairs from foils would be evidence that participants can learn based on object-specific regularities.

D.2.1 Methods

All participants were Cornell undergraduates who took part in exchange for course credit and provided informed consent consistent with the Cornell Institutional Review Board. The 22 participants in Experiment 1 (age: M = 20.7, SD = 1.75; 2 left handed; 10 female) were randomly and evenly assigned to either the typical or atypical conditions.

Familiarization

A statistically-structured familiarization sequence was presented, using PsyScope X B53 on a MacMini computer with a 17 in CRT monitor. Each picture was displayed for 300ms with a 700ms inter-stimulus interval (Brady & Oliva, 2008).

There were four categories of pictures: birds, dogs, fish, and flowers. For each category, four different exemplars were used (dog_1, dog_2, etc.). We employed both typical and atypical exemplars counterbalanced across participants. These four exemplars were selected based on pilot testing to be both recognizable for their semantic category and typicality or atypicality based on six candidate pictures. They were then grouped into eight pairs such that both the categories and the specific exemplars were predictive of picture order. For example, $bird_1$-dog_1 would always occur as a pair, as would $bird_2$-dog_2, $bird_3$-dog_3, and $bird_4$-dog_4 (Figure D.1). To control for any effect of specific pairings on learning, different categories and object pairings were employed across participants. Simply instructed to look at the pictures, participants saw each pair 28 times in random order without repetition.

Testing

After familiarization, participants were tested by comparing two pairs of pictures presented sequentially: 700 ms between pictures in the same pair, and 1200 ms separating the pairs. One pair was from the familiarization (e.g. $bird_1$-dog_1), and one was a foil pair (e.g. $bird_1$-dog_2; Figure D.3). The foils were designed to violate the structure only at the exemplar level, and not the category level. The participants were instructed to choose which of the pairs seemed more familiar, based on the familiarization task. No time constraint was imposed for their responses. There were 64 test trials. After the experiment, the participants completed a survey in which they rated the pictures they had seen on a scale of 1-5 for "interestingness" and typicality and were debriefed.

D.2.2 Results and Discussion

The current experiment was designed such that only object-specific knowledge could distinguish pairs seen during familiarization and foils. Performance was evaluated against chance (50%) for evidence of learning. Overall, participants reliably distinguished pairs from foils (M = 75.4%; SD = 22.8; $t(21) = 5.24$, $p < 0.0001$) indicating that participants learned object-specific regularities even though more abstract, category-level regularities were also present (Figure D.4).

We hypothesized that any contribution of categorical knowledge would be modulated by the typicality of the exemplars. To this end, we conducted a t-test examining the effects of typicality on test performance. We report no effect of exemplar typicality (Atypical M = 77.0%, SD = 21.5; Typical M = 73.9%, SD = 24.9, $t(21) = 0.315$, $p = 0.76$) suggesting that there was no contribution from

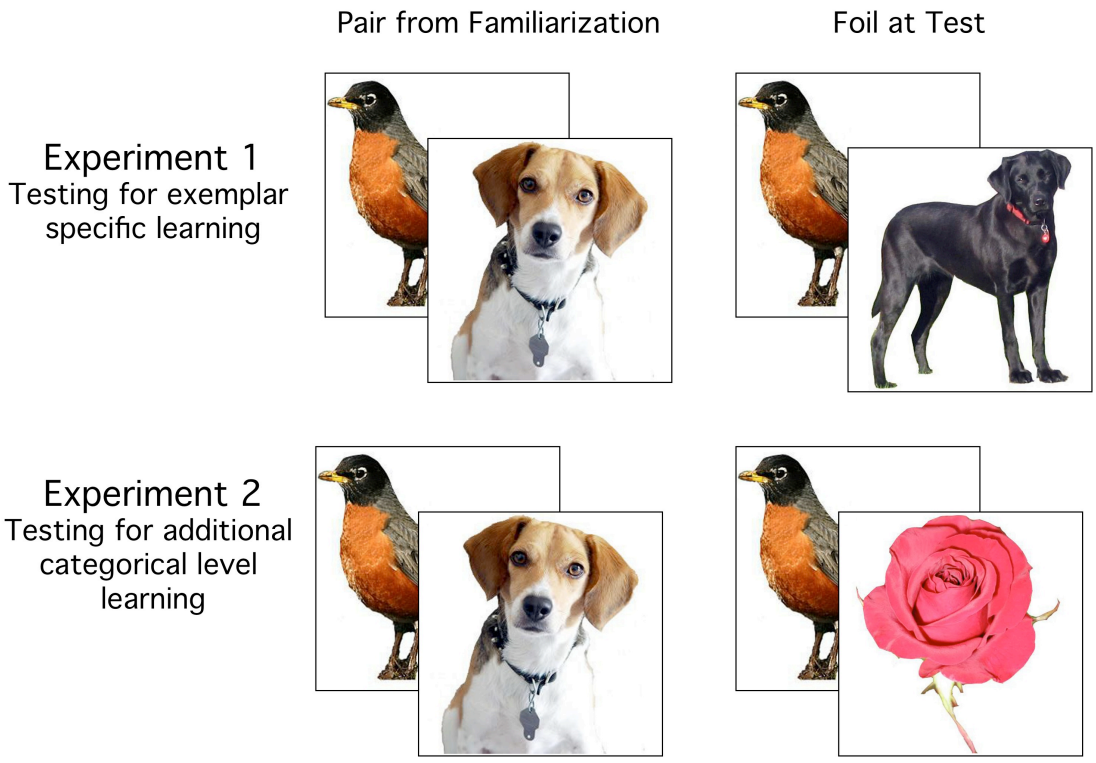


Figure D.3: The sole difference between experiments was the composition of foils used at test: In Experiment 1, foils were designed to assess learning at the object or exemplar-specific level. In this case, the category-level relationship of birds predicting dogs is held constant but the specific dog is changed to violate the object-specific regularities. In Experiment 2 foils allow for knowledge at both levels of abstraction (object and category) to influence test performance. Illustrated in the figure, now the birds predicting dogs regularity is being violated by the bird being presented with a flower in the foil.

category-level knowledge in the current experiment.

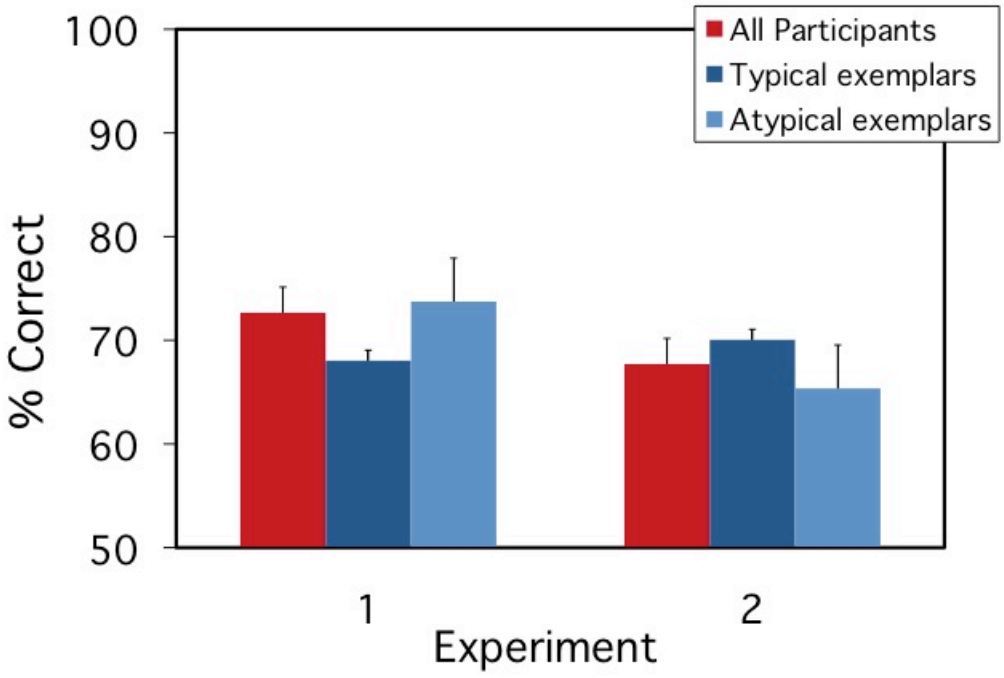


Figure D.4: Results for Experiments 1 and 2 are presented with the red bars representing performance for all participants and the blue bars presented means for participants grouped by whether they viewed typical or atypical exemplars. Error bars represent standard error of the mean.

D.3 Experiment 2: Testing for Additional Category-Level Knowledge

Having established that participants can learn from object-specific regularities in the presence of more abstract regularities, Experiment 2 examined whether learning occurs along multiple levels of abstraction simultaneously (e.g. objects: $bird_1$-dog_1; abstract, categories: birds-dogs). To this end, we changed the foils used at test so that both object-specific and category-level patterns could be used to distinguish pairs seen during familiarization from foils (e.g. $bird_1$-dog_1 vs. $bird_1$-$flower_3$; Figure D.3). We hypothesized that if participants learn from regularities at *both* levels of abstraction, test performance should increase in Ex-

periment 2 compared to Experiment 1 where only object-specific knowledge could be used at test. Conversely, failure to observe a significant increase in test performance suggests that learning does not occur at the abstract categorical level in addition to the object-specific knowledge demonstrated in Experiment 1.

Again, participants viewed objects that were either typical or atypical for their basic-level categories. In Experiment 1, we did not observe any asymmetry of performance between these groups; however, categorical knowledge would have interfered with test performance. In the current experiment, categorical knowledge would be of benefit. Thus, we hypothesize that if participants have access to category-level knowledge after familiarization, participants who view typical exemplars will have a greater boost in test performance than those who view atypical exemplars.

D.3.1 Methods

Another 24 participants were recruited from the same subject pool and randomly and evenly assigned to each condition (16 female, 1 left handed, age: M = 19.6, SD = 1.28, typical = 12). The procedure in this experiment differed from Experiment 1 in only one respect: the foil pairs during the test violated the statistical structure of the familiarization sequence at the exemplar and the category level (Figure D.3).

D.3.2 Results and Discussion

As with Experiment 1, we report significant learning overall in Experiment 2 compared to chance performance of 50% (M = 67.7%, SD = 21.2; $t(23) = 4.10$, $p < 0.0001$; Figure D.4). As with Experiment 1, we hypothesized that if categorical knowledge was acquired during exposure then it would be modulated by the typicality of exemplars. A t-test comparing performance based on object typicality revealed no effect of object typicality (atypical M = 65.4%, SD = 19.8; typical M = 70.1%, SD = 23.1; $t(23) = 0.53$, $p = 0.60$).

We also hypothesized that if participants learned from both category-level and object-specific regularities, Experiment 2 performance would increase compared to Experiment 1. Results from both experiments were analyzed in a two-way ANOVA (Experiment 1 vs. 2, Typicality). This analysis confirmed the pattern of results depicted in the bottom panel of Figure D.3: there is no main effect of Experiment ($F(1, 42) = 1.37$, $p = 0.25$). Additionally, we confirm that across both experiments there is no main effect of typicality of exemplars ($F(1, 42) = 0.01$, $p = 0.91$) and no interaction between these factors. Thus, test performance is equivalent across experiments indicating that participants likely did not acquire categorical knowledge during exposure to the familiarization stream.

All participants rated both typical and atypical pictures on "interestingness" and typicality. T-tests comparing ratings within categories revealed that participants rate atypical and typical exemplars distinctly and also rate the atypical exemplars as more interesting ($ts(150) > 5.0$; $ps < 0.001$). These results validate the assumption that participants view atypical and typical exemplars differently. Despite robust self-reported differences in typicality judgments, we consistently find no effect of exemplar typicality on learning.

Overall, these results suggest that violating both object-specific and category-level regularities learned during familiarization does not boost test performance compared to a test where only object-level regularities are violated. These findings cast doubt on the possibility that participants learn from predictive regularities at the higher level of category, in addition to the less abstract, object-specific patterns found consistently across both experiments.

D.4 General Discussion

We investigated statistical learning in a paradigm designed to reflect a key aspect of the complexity of daily experience: participants viewed streams of pictures with regularities at multiple levels of abstraction. Specifically, both individual objects and the categories to which they belonged predicted picture order, so that both object and categorical patterns could be learned. We consistently find evidence for learning at the lowest level of abstraction: participants respond at test according to the patterns of specific objects and do not show evidence of having learned at the more abstract level of categories even when abstract knowledge could aid test performance. Moreover, we find no modulation of learning by object typicality. These findings suggest that participants do not learn from more abstract regularities when less abstract, more grounded statistical information is present. Instead, they appear to be biased towards learning patterns based on specific objects or the least abstract pattern presented to them.

Previous research has shown that learners are sensitive to abstract, category-based regularities, but only under conditions where these are the only patterns

present and learnable.[1] In light of these findings, current results suggest that abstract patterns are more readily learnable when less abstract regularities are absent. When less abstract patterns are present, as in many daily experiences, our findings also suggest that learning likely starts based on the regularities of specific objects, only then proceeding towards more abstract patterns. Future research is required to more closely examine the time-course and the relationship between learning across levels of abstraction. Current work examines learning in adults based on known categories. Future work will seek to expand these results to younger learners and novel categories. Nevertheless, this finding may have important implications for more efficient teaching methods and could inform computational modeling of learning and development of human cognitive processes where the abstraction of representation is often an assumption built into the model.

Overall, this work aims to uncover how simple learning mechanisms operate in complex, naturalistic environments. We increased the complexity of the learning task, relative to previous experiments, by having environmental regularities at multiple levels of abstraction. These results inform the on-going debate as to whether domain-general learning mechanisms are largely unconstrained, as previous behavioral studies would have suggested. We believe that these results show some level of constraint on statistical learning where more grounded, less abstract statistical relationships are learned preferentially when categorical and object-specific knowledge are both learnable.

[1]See Experiment 4 in Brady and Oliva (2008) for an interesting demonstration: with less abstract regularities present, participants can learn the order of the labels for the categories of scene. It is not clear if this result is a demonstration of abstract-level learning when lower level regularities are present.

D.5 Acknowledgments

Drs. Dima Amso, Rick Dale and Jordan DeLong for helpful conversations, Claire Schmidt and Andrew Webb for data collection and in particular, we thank Dr. Michael Spivey for his support.

REFERENCES

Altmann, G., Dienes, Z., & Goode, A. (1995). Modality independence of implicitly learned grammatical knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 899-912.

Altmann, G., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science*, *33*(4), 583-609.

Amso, D., Davidson, M., Johnson, S., Glover, G., & Casey, B. (2005). Contributions of the hippocampus and the striatum to simple association and frequency-based learning. *Neuroimage*, *27*(2), 291–298.

Amso, D., Fitzgerald, M., Davidow, J., Gilhooly, T., & Tottenham, N. (2010). Visual exploration strategies and the development of infants' facial emotion discrimination. *Frontiers in Psychology*, *1*, 1-7.

Amso, D., & Johnson, S. (2006). Learning by selection: Visual search and object perception in young infants. *Developmental Psychology*, *42*(6), 1236-1245.

Arciuli, J., & Simpson, I. C. (2011). Statistical learning in typically developing children: the role of age and speed of stimulus presentation. *Developmental Science*, *14*(3), 464-473.

Aslin, R., Saffran, J., & Newport, E. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, *9*(4), 321–324.

Atallah, H., Frank, M., & O'Reilly, R. (2004). Hippocampus, cortex, and basal ganglia: Insights from computational models of complementary learning systems. *Neurobiology of Learning and Memory*, *82*(3), 253–267.

Baker, C., Olson, C., & Behrmann, M. (2004). Role of attention and perceptual grouping in visual statistical learning. *Psychological Science*, *15*(7), 460–466.

Balcetis, E., & Dunning, D. (2010). Wishful seeing: more desired objects are seen as closer. *Psychological Science*, *21*(1), 147-152.

Baldwin, D., Andersson, A., Saffran, J., & Meyer, M. (2008). Segmenting dynamic human action via statistical structure. *Cognition*, *106*(3), 1382-1407.

Bar, M., & Aminoff, E. (2003). Cortical analysis of visual context. *Neuron*, *38*(2), 347–358.

Barsalou, L., Simmons, W. K., Barbey, A., & Wilson, C. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences*, *7*(2), 84-91.

Bates, E., Bretherton, I., & Snyder, L. (1998). *From first words to grammar: Individual differences and dissociable mechanisms*. Cambridge, UK: Cambridge University Press.

Bavelier, D., Dye, M., & Hauser, P. (2006). Do deaf individuals see better? *Trends in Cognitive Sciences*, *10*(11), 512-518.

Baxter, M. G. (2009). Involvement of medial temporal lobe structures in memory and perception. *Neuron*, *61*(5), 667-677.

Beckman, M., & Edwards, J. (2000). The ontogeny of phonological categories and the primacy of lexical learning in linguistic development. *Child Development*, *71*(1), 240-249.

Berridge, K. C. (2000). Reward learning: Reinforcement, incentives, and expectations. *Psychology of Learning and Motivation*, *40*, 223-278.

Bhatt, R., & Quinn, P. (2011). How does learning impact development in infancy? the case of perceptual organization. *Infancy*, *16*(1), 2-38.

Binkofski, F., Buccino, G., Posse, S., Seitz, R., Rizzolatti, G., & Freund, H. (1999). A fronto-parietal circuit for object manipulation in man: evidence from an fmri-study. *European Journal of Neuroscience*, *11*(9), 3276-3286.

Blake, R., & Logothetis, N. (2002). Visual completion. *Nature Reviews Neuroscience*, *3*(1), 13-21.

Booth, M., & Rolls, E. (1998). View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. *Cerebral Cortex*, *8*(6), 510-523.

Brady, T., & Oliva, A. (2008). Statistical learning using real-world scenes: Extracting categorical regularities without conscious intent. *Psychological Science*, *19*(7), 678–685.

Bregman, A. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: The MIT Press.

Broadbent, D. E. (1958). *Perception and communication*. Oxford, UK: Oxford University Press.

Bruner, J. (1957). On perceptual readiness. *Psychological Review*, *64*(2), 123-152.

Buckley, M., Booth, M., Rolls, E., & Gaffan, D. (2001). Selective perceptual impairments after perirhinal cortex ablation. *Journal of Neuroscience*, *21*(24), 9824-9836.

Burr, D., Morrone, M., & Ross, J. (1994). Selective suppression of the magnocellular visual pathway during saccadic eye movements. *Nature*, *371*(6497), 511-513.

Cabeza, R., Ciaramelli, E., Olson, I., & Moscovitch, M. (2008). The parietal cortex and episodic memory: an attentional account. *Nature Reviews Neuroscience*, *9*(8), 613-625.

Cabeza, R., Mazuz, Y. S., Stokes, J., Kragel, J. E., Woldorff, M. G., Ciaramelli, E., et al. (2011). Overlapping parietal activity in memory and perception: Evidence for the attention to memory model. *Journal of Cognitive Neuroscience*, *23*(11), 3209-3217.

Caruso, E., Mead, N., & Balcetis, E. (2009). Political partisanship influences perception of biracial candidates' skin tone. *Proceedings of the National*

Academy of Sciences, *106*(48), 20168-20173.

Catena, J., Scholl, B., Isola, P., & Turk-Browne, N. (2010). The onset and offset of visual statistical learning. *Manuscript submitted for publication*.

Census Bureau, U. S. (2007a). *State & county quickfacts: Ithaca, N.Y.* Retrieved September 19th, 2011, from `http://quickfacts.census.gov`

Census Bureau, U. S. (2007b). *State & county quickfacts: Syracuse, N.Y.* . Retrieved September 19th, 2011, from `http://quickfacts.census.gov`

Charness, N., Reingold, E., Pomplun, M., & Stampe, D. (2001). The perceptual aspect of skilled performance in chess: Evidence from eye movements. *Psychological science*, *29*(8), 1146-1152.

Chen, Y., Repp, B., & Patel, A. (2002). Spectral decomposition of variability in synchronization and continuation tapping: Comparisons between auditory and visual pacing and feedback conditions. *Human Movement Science*, *21*, 515-532.

Chomsky, N. (1959). A review of bf skinner's verbal behavior. *Language*, *35*(1), 26-58.

Chomsky, N. (1967). Recent contributions to the theory of innate ideas. *Synthese*, *17*(1), 2-11.

Christiansen, M., Onnis, L., & Hockema, S. (2009). The secret is in the sound: From unsegmented speech to lexical categories. *Developmental Science*, *12*(3), 388-395.

Chun, M., & Phelps, E. (1999). Memory deficits for implicit contextual information in amnesic subjects with hippocampal damage. *Nature Neuroscience*, *2*, 844–847.

Clark, D., & Ivry, R. (2010). Multiple systems for motor skill learning. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(4), 461-467.

Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, *108*, 804-809.

Cleland, T. A. (2010). Early transformations in odor representation. *Trends in Neurosciences*, *33*(3), 130-139.

Clopper, C., & Pisoni, D. (2004). Effects of talker variability on perceptual learning of dialects. *Language and Speech*, *47*(3), 207-239.

Cohen, L., Atkinson, D., & Chaput, H. (2000). Habit 2000: A new program for testing infant perception and cognition. *[Computer software]. Austin, the University of Texas*.

Conway, C., Bauernschmidt, A., Huang, S., & Pisoni, D. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition*, *114*(3), 356-371.

Conway, C., & Christiansen, M. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology-Learning Memory and Cognition*, *31*(1), 24–38.

Conway, C., & Christiansen, M. (2006). Statistical learning within and between modalities: pitting abstract against stimulus-specific representations. *Psychological Science*, *17*(10), 905-912.

Conway, C., & Christiansen, M. (2009). Seeing and hearing in space and time: Effects of modality and presentation rate on implicit statistical learning. *European Journal of Cognitive Psychology*, *21*, 561-580.

Cox, R. (1996). Afni: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, *29*(3), 162-173.

Creel, S., Newport, E., & Aslin, R. (2004). Distant melodies: Statistical learning

of nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(5), 1119-1130.

Curran, T. (1997). Higher-order associative learning in amnesia: Evidence from the serial reaction time task. *Journal of Cognitive Neuroscience*, *9*(4), 522–533.

Dale, R., Kehoe, C., & Spivey, M. (2007). Graded motor responses in the time course of categorizing atypical exemplars. *Memory & cognition*, *35*(1), 15-28.

Daw, N., Courville, A., & Touretzky, D. (2002). Dopamine and inference about timing. In *Proceedings of the 2nd international conference on development and learning* (p. 271-276). Cambridge, MA.

Daw, N., Courville, A., & Touretzky, D. (2006). Representation and timing in theories of the dopamine system. *Neural computation*, *18*(7), 1637-1677.

Deco, G., & Rolls, E. T. (2004). A neurodynamical cortical model of visual attention and invariant object recognition. *Vision Research*, *44*(6), 621-642.

De Haan, M., & Nelson, C. (1999). Brain activity differentiates face and object processing in 6-month-old infants. *Developmental Psychology*, *35*(4), 1113-1121.

den Ouden, H. E. M., Daunizeau, J., Roiser, J., Friston, K., & Stephan, K. (2010). Striatal prediction error modulates cortical coupling. *Journal of Neuroscience*, *30*(9), 3210-3219.

den Ouden, H. E. M., Friston, K. J., Daw, N., McIntosh, A. R., & Stephan, K. E. (2009). A Dual-Role for Prediction Error in Associative Learning. *Cerebral Cortex*, *19*, 1175–1851.

Desimone, R. (1991). Face-selective cells in the temporal cortex of monkeys. *Journal of Cognitive Neuroscience*, *3*(1), 1-8.

Dommett, E., Coizet, V., Blaha, C., Martindale, J., Lefebvre, V., Walton, N., et al.

(2005). How visual stimuli activate dopaminergic neurons at short latency. *Science*, *307*(5714), 1476-1479.

Egner, T., Monti, J., & Summerfield, C. (2010). Expectation and surprise determine neural population responses in the ventral visual stream. *Journal of Neuroscience*, *30*(49), 16601-16608.

Eichenbaum, H., & Cohen, N. (2001). *From conditioning to conscious recollection: Memory systems of the brain*. USA: Oxford University Press.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.

Emberson, L. L., & Amso, D. (in preparation). The medial temporal lobe supports changes in object perception.

Emberson, L. L., Conway, C., & Christiansen, M. (2011). Changes in presentation rate have opposite effects timing is everything: Changes in presentation rate have opposite effects on auditory and visual implicit statistical learning learning. *Quarterly Journal of Experimental Psychology*, *64*(5), 1021-1040.

Emberson, L. L., Liu, R., & Zevin, J. D. (under review). Is statistical learning constrained by lower level perceptual organization? *Cognition*.

Emberson, L. L., Misyak, J. B., Schwade, J., Christiansen, M., & Goldstein, M. (in preparation). How abstract is statistical learning? comparing learning in visual and auditory perceptual modalities in infancy.

Emberson, L. L., & Rubinstein, D. (under review). Staying grounded: Is statisitcal learning constrained to less abstract patterns?

Feldman, N., Griffiths, T., & Morgan, J. (2009). The influence of categories on perception: Explaining the perceptual magnet effect as optimal statistical

inference. *Psychological Review*, *116*(4), 752-782.

Fernald, A., & Hurtado, N. (2006). Names in frames: Infants interpret words in sentence frames faster than words in isolation. *Developmental Science*, *9*(3), 33–40.

Fernald, A., Swingley, D., & Pinto, J. (2001). When half a word is enough: Infants can recognize spoken words using partial phonetic information. *Child Development*, *72*(4), 1003-1015.

Fiser, J., & Aslin, R. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological science*, *12*(6), 499–504.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences*, *99*(24), 15822-15826.

Flynn, M., & Dowell, R. (1999). Speech perception in a communicative context: An investigation using question/answer pairs. *Journal of Speech, Language, and Hearing Research*, *42*(3), 540-552.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Fried, I., MacDonald, K., & Wilson, C. (1997). Single neuron activity in human hippocampus and amygdala during recognition of faces and objects. *Neuron*, *18*(5), 753-765.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *360*(1456), 815-836.

Gauthier, I., & Tarr, M. (1997). Becoming a "greeble" expert: Exploring mechanisms of face recognition. *Vision Research*, *37*(12), 1673–1682.

Gauthier, I., Tarr, M., Anderson, A., Skudlarski, P., & Gore, J. (1999). Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nature Neuroscience*, *2*(6), 568-573.

Gebhart, A., Aslin, R., & Newport, E. (2009). Changing structures in midstream: Learning along the statistical garden path. *Cognitive Science*, *33*(6), 1087-1116.

Ghetti, S., & Angelini, L. (2008). The development of recollection and familiarity in childhood and adolescence: Evidence from the dual-process signal detection model. *Child Development*, *79*(2), 339-358.

Gibson, E. J. (1969). *Principals of perceptual learning and development*. New York: Appleton-Century-Crofts.

Gibson, E. J., & Pick, A. . (2000). *An ecological approach to perceptual learning and development*. Oxford: Oxford University Press.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. London: George Allen and Unwin Ltd.

Gibson, J. J., & Gibson, E. J. (1955). Perceptual learning: Differentiation or enrichment? *Psychological Review*, *62*(1), 32-41.

Glenberg, A. M. (1997). What memory is for. *Brain and Behavioral Sciences*, *20*, 1-55.

Goldstein, M., & Schwade, J. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science*, *19*(5), 515–523.

Goldstein, M., Waterfall, H., Lotem, A., Halpern, J., Schwade, J., Onnis, L., et al. (2010). General cognitive principles for learning structure in time and space. *Trends in Cognitive Sciences*, *14*(6), 249-258.

Goldwater, S., Griffiths, T., & Johnson, M. (2009). A bayesian framework for word segmentation: Exploring the effects of context. *Cognition*, *112*(1), 21-54.

Gómez, R. (2006). Attention & performance xxi: Processes of change in brain and cognitive development. In Y. Munakata & M. Johns (Eds.), (p. 87-110).

Oxford, UK: Oxford University Press.

Gomez, R., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, *70*(2), 109–135.

Gopnik, A., & Tenenbaum, J. (2007). Bayesian networks, bayesian learning and cognitive development. *Developmental Science*, *10*(3), 281-287.

Graf, P., & Schacter, D. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(3), 501-518.

Graf Estes, K., Evans, J., Alibali, M., & Saffran, J. (2007). Can Infants Map Meaning to Newly Segmented Words?: Statistical Segmentation and Word Learning. *Psychological Science*, *18*(3), 254–260.

Grant, E., & Spivey, M. (2003). Eye movements and problem solving. *Psychological science*, *14*(5), 464-466.

Green, C., & Bavelier, D. (2008). Exercising your brain: A review of human brain plasticity and training-induced learning. *Psychology and Aging*, *23*(4), 692-701.

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Sciences*, *10*(1), 14-23.

Guttman, N., & Kalish, H. (1956). Discriminability and stimulus generalization. *Journal of Experimental Psychology*, *51*(1), 79-88.

Haan, M. de, & Nelson, C. (1997). Recognition of the mother. *Child Development*, *68*(2), 187-210.

Haith, M., Hazan, C., & Goodman, G. (1988). Expectation and anticipation of dynamic visual events by 3.5-month-old babies. *Child Development*, *59*(2), 467-479.

Handel, S., Weaver, M., & Lawson, G. (1983). Effect of rhythmic grouping on stream segregation. *Journal of Experimental Psychology: Human Perception and Performance*, *9*(4), 637-651.

Hannula, D., Ryan, J., Tranel, D., & Cohen, N. (2007). Rapid onset relational memory effects are evident in eye movement behavior, but not in hippocampal amnesia. *Journal of Cognitive Neuroscience*, *19*(10), 1690-1705.

Harel, A., Gilaie-Dotan, S., Malach, R., & Bentin, S. (2010). Top-down engagement modulates the neural expressions of visual expertise. *Cerebral Cortex*, *10*(2304–2318).

Harris, I., Benito, C., Ruzzoli, M., & Miniussi, C. (2008). Effects of right parietal transcranial magnetic stimulation on object identification and orientation judgments. *Journal of Cognitive Neuroscience*, *20*(5), 916-926.

Hasselmo, M., Rolls, E., Baylis, G., & Nalwa, V. (1989). Object-centered encoding by face-selective neurons in the cortex in the superior temporal sulcus of the monkey. *Experimental Brain Research*, *75*(2), 417-429.

Haxby, J., Gobbini, M., Furey, M., Ishai, A., Schouten, J., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*(5539), 2425-2430.

Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6-12. *Journal of Phonetics*, *28*(4), 377-396.

Heisz, J., & Shore, D. (2008). More efficient scanning for familiar faces. *Journal of Vision*, *8*(1), 1-10.

Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Advances in Infancy Research*, *5*, 69-95.

Ison, M., & Quiroga, R. (2008). Selectivity and invariance for visual object perception. *Frontiers in Bioscience*, *1*, 4889-4903.

Ito, M., Tamura, H., Fujita, I., & Tanaka, K. (1995). Size and position invariance of neuronal responses in monkey inferotemporal cortex. *Journal of Neurophysiology*, *73*(1), 218-226.

Jemel, B., Pisani, M., Calabria, M., Crommelinck, M., & Bruyer, R. (2003). Is the n170 for faces cognitively penetrable? evidence from repetition priming of mooney faces of familiar and unfamiliar persons. *Cognitive Brain Research*, *17*(2), 431-446.

Johnson, E., Westrek, E., Nazzi, T., & Cutler, A. (2011). Infant ability to tell voices apart rests on language experience. *Developmental Science*, *14*(5), 1002-1011.

Johnson, M., Posner, M., & Rothbart, M. (1994). Facilitation of saccades toward a covertly attended location in early infancy. *Psychological Science*, *5*(2), 90-93.

Johnson, S. (Ed.). (2010). *Neoconstructivism: The new science of cognitive development*. Oxford, UK: Oxford University Press.

Johnson, S., Slemmer, J., & Amso, D. (2004). Where infants look determines how they see: Eye movements and object perception performance in 3-month-olds. *Infancy*, *6*(2), 185-201.

Jolicoeur, P., Gluck, M., & Kosslyn, S. (1984). Pictures and names: Making the connection. *Cognitive Psychology*, *16*(2), 243-275.

Jungé, J., Scholl, B., & Chun, M. (2007). How is spatial context learning integrated over signal versus noise? A primacy effect in contextual cueing. *Visual cognition*, *15*(1), 1–11.

Jusczyk, P., & Aslin, R. (1995). Infants detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, *29*(1), 1-23.

Kanwisher, N., McDermott, J., & Chun, M. (1997). The fusiform face area: a

module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*(11), 4302-4311.

Keele, S., Ivry, R., Mayr, U., Hazeltine, E., & Heuer, H. (2003). The cognitive and neural architecture of sequence representation. *Psychological Review*, *110*(2), 316–339.

Kelly, D., Quinn, P., Slater, A., Lee, K., Ge, L., & Pascalis, O. (2007). The other-race effect develops during infancy. *Psychological Science*, *18*(12), 1084-1089.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as bayesian inference. *Annual review of psychology*, *55*, 271-304.

Kirkham, N., Slemmer, J., & Johnson, S. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35-B42.

Kobatake, E., & Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *Journal of Neurophysiology*, *71*(3), 856-867.

Kóbor, I., Furedi, L., Kovács, G., Spence, C., & Vidnyánszky, Z. (2006). Back-to-front: Improved tactile discrimination performance in the space you cannot see. *Neuroscience Letters*, *400*(1-2), 163-167.

Kreiman, G., Koch, C., & Fried, I. (2000). Category-specific visual responses of single neurons in the human medial temporal lobe. *Nature Neuroscience*, *3*(9), 946-953.

Kubovy, B. (1988). Should we resist the seductiveness of the space:time::vision:audition analogy? *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 318-320.

Kubovy, M., Holcombe, A. O., & Wagemans, J. (1998). On the lawfulness of

grouping by proximity. *Cognitive Psychology*, *35*(71-98).

Kuhl, P., & Rivera-Gaxiola, M. (2008). Neural substrates of language acquisition. *Annual Review of Neuroscience*, *31*, 511-534.

Kuhl, P., Williams, K., Lacerda, F., Stevens, K., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*(5044), 606-608.

Lany, J., & Saffran, J. (2010). From statistics to meaning. *Psychological science*, *21*(2), 284-291.

Leech, R., Holt, L., Devlin, J., & Dick, F. (2009). Expertise with artificial nonspeech sounds recruits speech-sensitive cortical regions. *Journal of Neuroscience*, *29*(16), 5234-5239.

Leopold, D., & Logothetis, N. (1999). Multistable phenomena: changing views in perception. *Trends in Cognitive Sciences*, *3*(7), 254-264.

Lewis, M. (1969). Infants' responses to facial stimuli during the first year of life. *Developmental Psychology*, *1*(2), 75-86.

Lewkowicz, D., & Ghazanfar, A. (2009). The emergence of multisensory systems through perceptual narrowing. *Trends in cognitive sciences*, *13*(11), 470-478.

Lew-Williams, C., & Fernald, A. (2007). Young children learning spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, *18*(3), 193-198.

Little, A., Lipsitt, L., & Rovee-Collier, C. (1984). Classical conditioning and retention of the infant's eyelid response: Effects of age and interstimulus interval. *Journal of experimental child psychology*, *37*(3), 512-524.

Liu, R., & Holt, L. (2011). Neural changes associated with nonspeech auditory category learning parallel those of speech category acquisition. *Journal of Cognitive Neuroscience*, *23*(3), 683-698.

Logothetis, N., & Sheinberg, D. (1996). Visual object recognition. *Annual Review of Neuroscience*, *19*(1), 577-621.

Lotto, A. J. (2000). Language acquisition as complex category formation. *Phonetica*, *57*(2-4), 189-196.

Lotto, A. J., & Holt, L. L. (2000). The illusion of the phoneme. In S. J. Billings, J. P. Boyle, & A. M. Griffith (Eds.), *Chicago linguistic society, volume 35: The panels.* (p. 191-204). Chicago: Chicago Linguistic Society.

MacWhinney, B. (1991). *The childes project: Tools for analyzing talk*. NJ: Lawrence Erlbaum Associates, Inc.

Mahar, D., Mackenzie, B., & McNicol, D. (1994). Modality-specific differences in the processing of spatially, temporally, and spatiotemporally distri- buted information. *Perception*, *23*, 1369-1386.

Marcus, G., Vijayan, S., Rao, S., & Vishton, P. (1999). Rule learning by seven-month-old infants. *Science*, *283*(5398), 77.

Mareschal, D., Johnson, M., Sirois, S., Spratling, M. W., Thomas, M. S. C., & Westermann, G. (2007). *Neuroconstructivism: How the brain constructs cognition*. Oxford, UK: Oxford University Press.

Markman, A., & Brendl, C. (2005). Constraining theories of embodied cognition. *Psychological Science*, *16*(1), 6-10.

Mates, J., Müller, U., Radil, T., & Pöppel, E. (1994). Temporal integration in sensorimotor synchronization. *Journal of Cognitive Neuroscience*, *6*(4), 332-340.

Maye, J., Weiss, D., & Aslin, R. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science*, *11*(1), 122-134.

Maye, J., Werker, J., & Gerken, L. (2002). Infant sensitivity to distributional

information can affect phonetic discrimination. *Cognition*, *82*(3), 101–111.

McClelland, J. (2002). Prediction-error driven learning: the engine of change in cognitive development. In *Proceedings of the 2nd international conference on development and learning* (p. 43).

McClelland, J., McNaughton, B., & O'Reilly, R. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological review*, *102*(3), 419–457.

McClure, S., Berns, G., & Montague, P. (2003). Temporal prediction errors in a passive learning task activate human striatum. *Neuron*, *38*(2), 339–346.

McGowan, R., Nittrouer, S., & Manning, C. (2004). Development of [invr] in young, midwestern, american children. *Journal of Acoustical Society of America*, *115*, 871-884.

McMullen, E., & Saffran, J. (2004). Music and language: A developmental comparison. *Music Perception*, *21*(3), 289-311.

McNaughton, B., & Nadel, L. (1989). *Neuroscience and connectionist theory*. NJ: Lawrence Erlbaum Associates.

McNealy, K., Mazziotta, J. C., & Dapretto, M. (2006). Cracking the language code: Neural mechanisms underlying speech parsing. *Journal of Neuroscience*, *26*(29), 7629-7639.

McNealy, K., Mazziotta, J. C., & Dapretto, M. (2010). The neural basis of speech parsing in children and adults. *Developmental Science*, *13*(385-406).

McNealy, K., Mazziotta, J. C., & Dapretto, M. (in press). Age and experience shape developmental changes in the neural basis of language-related learning. *Developmental Science*.

Milgram, N., MacLeod, C., & Petit, T. (1987). Neuroplasticity, learning, and

memory. In N. Milgram, C. MacLeod, & T. Petit (Eds.), *Neuroplasticity, learning, and memory: Proceedings of a Symposium held at the University of Toronto, Scarborough, Ontario, March 25, 1986* (p. 1-16). New York: Alan R. Liss Inc.

Miller, G. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81-97.

Miller, R., Barnet, R. C., & Grahame, N. J. (1995). Assessment of the rescorla-wagner model. *Psychological bulletin*, *117*(3), 363-386.

Monosov, I., Sheinberg, D., & Thompson, K. (2010). Paired neuron recordings in the prefrontal and inferotemporal cortices reveal that spatial selection precedes object identification during visual search. *Proceedings of the National Academy of Sciences*, *107*(29), 13105-13110.

Mooney, C. M. (1957). Age in the development of closer ability in children. *Canadian journal of Psychology*, *11*(4), 219-226.

Narayan, C., Werker, J., & Beddor, P. (2010). The interaction between acoustic salience and language experience in developmental speech perception: Evidence from nasal place discrimination. *Developmental Science*, *13*(3), 407-420.

Needham, A., Dueker, G., & Lockhead, G. (2005). Infants' formation and use of categories to segregate objects. *Cognition*, *94*(5), 215-240.

Needham, A., & Modi, A. (1999).) infants' use of prior experiences with objects in object segregation: Implications for object recognition in infancy. *Advances in child development and behavior*, *27*, 99-133.

Neisser, U. (1967). *Cognitive psychology*. New York, NY: Appleton-Century-Crofts.

Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive*

psychology. New York: WH Freeman & Co.

Nelson, C. A. (2010). The development and neural bases of face recognition. *Infant and Child Development*, *10*(1-2), 3-18.

Neville, H. (2006). Different profiles of plasticity within human cognition. In Y. Munakata & M. Johnson (Eds.), *Processes of change in brain and cognitive development: Attention and performance xxi. different profiles of plasticity within human cognition* (p. 287-314). Oxford, UK: Oxford University Press.

Neville, H., & Bavelier, D. (2002). Human brain plasticity: evidence from sensory deprivation and altered language experience. *Progress in Brain Research*, *138*, 177-188.

Nishina, S., Seitz, A., Kawato, M., & Watanabe, T. (2007). Effect of spatial distance to the task stimulus on task-irrelevant perceptual learning of static gabors. *Journal of Vision*, *7*(13), 1-10.

Nissen, M., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*(1), 1-32.

Nomura, E., & Reber, P. (2008). A review of medial temporal lobe and caudate contributions to visual category learning. *Neuroscience and Biobehavioral Reviews*, *32*(2), 279–291.

Nusbaum, H. C., & Henly, A. S. (1992). Listening to speech through an adaptive window of analysis. In M. E. H. Schouten (Ed.), *The auditory processing of speech: from sounds to words* (p. 339-348). Berlin: Mouton de Gruyter.

Oakes, L., & Cohen, L. (1990). Infant perception of a causal event. *Cognitive Development*, *5*(2), 193-207.

Oakes, L., & Cohen, L. (1995). Infant causal perception. *Advances in Infancy Research*, *9*, 1-54.

Obleser, J., & Kotz, S. (2010). Expectancy constraints in degraded speech modu-

late the language comprehension network. *Cerebral Cortex*, *20*(3), 633-640.

Onnis, L., Monaghan, P., Christiansen, M., & Chater, N. (2004). Variability is the spice of learning, and a crucial ingredient for detecting and generalizing in nonadjacent dependencies. In *Proceedings of the 26th annual conference of the cognitive science society.* Hillsdale, NJ:: Lawrence Erlbaum Associates.

Onnis, L., Waterfall, H., & Edelman, S. (2008). Learn locally, act globally: Learning language from variation set cues. *Cognition*, *109*(3), 423–430.

Op De Beeck, H., & Vogels, R. (2000). Spatial sensitivity of macaque inferior temporal neurons. *The Journal of Comparative Neurology*, *426*(4), 505-518.

Orbán, G., Fiser, J., Aslin, R., & Lengyel, M. (2008). Bayesian learning of visual chunks by human observers. *Proceedings of the National Academy of Sciences*, *105*(7), 2745-2750.

O'Reilly, R. (2006). Biologically based computational models of high-level cognition. *Science*, *314*(5796), 91–94.

Pacton, S., & Perruchet, P. (2008). An attention-based associative account of adjacent and nonadjacent dependency learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(1), 80-96.

Pagel, B., Heed, T., & Röder, B. (2009). Change of reference frame for tactile localization during child development. *Developmental Science*, *12*(6), 929-937.

Pagnoni, G., Zink, C., Montague, P., & Berns, G. (2002). Activity in human ventral striatum locked to errors of reward prediction. *Nature Neuroscience*, *5*(2), 97-98.

Palmer, S., & Rock, I. (1994). Rethinking perceptual organization: The role of uniform connectedness. *Psychological bulletin & review*, *1*(1), 29-55.

Pascalis, O., Scott, L., Kelly, D., Shannon, R., Nicholson, E., Coleman, M., et al.

(2005). Plasticity of face processing in infancy. *Proceedings of the National Academy of Sciences*, *102*(14), 5297-5300.

Pelucchi, B., Hay, J., & Saffran, J. (2009). Statistical learning in a natural language by 8-month-old infants. *Child Development*, *80*(3), 674-685.

Penney, C. G. (1989). Modality effects and the structure of short-term verbal memory. *Memory & Cognition*, *17*, 398-422.

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, *10*(5), 233-238.

Peterson, G., & Barney, H. (1952). Control methods used in a study of the vowels. *Journal of Acoustical Society of America*, *24*(2), 175-184.

Piaget, J. (1971). *Biology and knowledge*. Chicago: University of Chicago Press.

Pinker, S. (1995). *The language instinct*. New York, NY: Harper Perennial.

Poeppel, D., Idsardi, W., & Van Wassenhove, V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *363*(1493), 1071-1086.

Poldrack, R., & Packard, M. (2003). Competition among multiple memory systems: converging evidence from animal and human brain studies. *Neuropsychologia*, *41*(3), 245–251.

Pomplun, M., Ritter, H., & Velichkovsky, B. (1996). Disambiguating complex visual information: Towards communication of personal views of a scene. *Perception*, *25*(8), 931-948.

Pons, F., Lewkowicz, D., Soto-Faraco, S., & Sebastián-Gallés, N. (2009). Narrowing of intersensory speech perception in infancy. *Proceedings of the National Academy of Sciences*, *106*(26), 10598–10602.

Port, R. (2007). How are words stored in memory? beyond phones and phonemes. *New Ideas in Psychology*, *25*(2), 143-170.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, *2*(509-522).

Pylyshyn, Z. (1999). Is vision continuous with cognition? the case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, *22*(3), 341-365.

Quiroga, R., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*, *435*(7045), 1102-1107.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, *124*(3), 372-422.

Reber, A. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, *118*(3), 219-235.

Reber, A., Kassin, S., Lewis, S., & Cantor, G. (1980). On the relationship between implicit and explicit modes in the learning of a complex rule structure. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(5), 492-502.

Redgrave, P., & Gurney, K. (2006). The short-latency dopamine signal: a role in discovering novel actions? *Nature Reviews Neuroscience*, *7*(12), 967–975.

Redgrave, P., Prescott, T., & Gurney, K. (1999). Is the short-latency dopamine response too short to signal reward error? *Trends in Neurosciences*, *22*(4), 146-151.

Remez, R., Rubin, P., Pisoni, D., & Carrell, T. (1981). Speech perception without traditional speech cues. *Science*, *212*(4497), 947-949.

Rescorla, R., & Wagner, A. (1972). A theory of pavolvian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In A. Black & W. Prokasy (Eds.), *Classical conditioning ii: Current research and*

theory (p. 64-99). New York: Appleton-Century-Crofts.

Richards, J., & Gibson, T. (1997). Extended visual fixation in young infants: Look distributions, heart rate changes, and attention. *Child Development*, *68*(6), 1041-1056.

Richmond, J., & Nelson, C. (2009). Relational memory during infancy: evidence from eye tracking. *Developmental Science*, *12*(4), 549-556.

Robinson, C. W., & Sloutsky, V. M. (2007). Visual statistical learning: Getting some help from the auditory modality. In D. S. M. . J. G. Trafton (Ed.), *Proceedings of the 29th annual cognitive science society* (p. 611-616). Austin, TX: Cognitive Science Society.

Rodríguez-Fornells, A., Cunillera, T., Mestres-Missé, A., & Diego-Balaguer, R. de. (2009). Neurophysiological mechanisms involved in language learning in adults. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*(1536), 3711-3735.

Rolls, E. T. (1992). Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *335*(1273), 11-20.

Rolls, E. T. (2000). Functions of the primate temporal review lobe cortical visual areas in invariant visual object and face recognition. *Neuron*, *27*(2), 205-218.

Romberg, A., & Saffran, J. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(6), 906-914.

Rosch, E., Mervis, C., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive psychology*, *8*(3), 382-439.

Rose, S., Gottfried, A., Melloy-Carminar, P., & Bridger, W. (1982). Familiarity

and novelty preferences in infant recognition memory: Implications for information processing. *Developmental Psychology*, *18*(5), 704-713.

Rost, G., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, *12*(2), 339-349.

Rovee-Collier, C. (1986). The rise and fall of infant classical conditioning research: Its promise for the study of early development. *Advances in Infancy Research*, *4*, 139-159.

Rovee-Collier, C., & Cuevas, K. (2009). Multiple memory systems are unnecessary to account for infant memory development: An ecological model. *Developmental Psychology*, *45*(1), 160-174.

Saffran, J. (2002). Constraints on statistical language learning. *Journal of Memory and Language*, *47*(1), 172–196.

Saffran, J., Aslin, R., & Newport, E. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926.

Saffran, J., Johnson, E., Aslin, R., & Newport, E. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52.

Saffran, J., Newport, E., Aslin, R., Tunick, R., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science*, *8*(2), 101-105.

Saffran, J., Pollak, S., Seibel, R., & Shkolnik, A. (2007). Dog is a dog is a dog: Infant rule learning is not specific to language. *Cognition*, *105*(3), 669–680.

Saffran, J., & Sahni, S. D. (in press). Learning the sounds of language. In M. Joanisse, M. Spivey, & M. K (Eds.), *Cambridge handbook of psycholinguistics.* Cambridge: Cambridge University Press.

Saffran, J., & Thiessen, E. (2003). Pattern induction by infant language learners. *Developmental Psychology*, *39*(3), 484-494.

Saffran, J., & Thiessen, E. (2007). Domain-general learning capacities. In E. Hoff & M. Shatz (Eds.), *Blackwell handbook of language development* (p. 68-86). New Jersey: Wiley-Blackwell.

Sakata, H., & Kusunoki, M. (1992). Organization of space perception: Neural representation of three-dimensional space in the posterior parietal cortex. *Current Opinion in Neurobiology*, *2*(2), 170-174.

Sakata, H., Taira, M., Kusunoki, M., Murata, A., Tsutsui, K., Tanaka, Y., et al. (1999). Neural representation of three-dimensional features of manipulation objects with stereopsis. *Experimental Brain Research*, *128*(1), 160-169.

Schacter, D. L., & Tulving, E. (1994). *Memory systems*. Cambridge, MA: MIT Press.

Schendan, H., Searl, M., Melrose, R., & Stern, C. (2003). An fMRI study of the role of the medial temporal lobe in implicit and explicit sequence learning. *Neuron*, *37*(6), 1013–1025.

Schlesinger, M., Amso, D., & Johnson, S. (2007). The neural basis for visual selective attention in young infants: A computational account. *Adaptive Behavior*, *15*(2), 135-148.

Schultz, W. (2006). Behavioral theories and the neurophysiology of reward. *Annual review of psychology*, *57*, 87–115.

Schultz, W., Dayan, P., & Montague, P. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593-1599.

Seidenberg, M., MacDonald, M., & Saffran, J. (2002). Does grammar start where statistics stop? *Science*, *298*(5593), 553-554.

Serniclaes, W., Heghe, S., Mousty, P., Carré, R., & Sprenger-Charolles, L. (2004). Allophonic mode of speech perception in dyslexia. *Journal of experimental child psychology*, *87*(4), 336-361.

Shanks, D., Johnstone, T., & Staggs, L. (1997). Abstraction processes in artificial grammar learning. *The Quarterly Journal of Experimental Psychology A*, *50*(1), 216-252.

Shankweiler, D., Strange, W., & Verbrugge, R. (1977). Speech and the problem of perceptual constancy. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting, knowing: Toward an ecological psychology* (p. 315-345). NJ: Erlbaum.

Shannon, R., Zeng, F., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, *270*(5234), 303-304.

Shikata, E., Hamzei, F., Glauche, V., Knab, R., Dettmers, C., Weiller, C., et al. (2001). Surface orientation discrimination activates caudal and anterior intraparietal sulcus in humans: an event-related fmri study. *Journal of Neurophysiology*, *85*(3), 1309-1314.

Shohamy, D., Myers, C., Grossman, S., Sage, J., Gluck, M., & Poldrack, R. (2004). Cortico-striatal contributions to feedback-based learning: converging data from neuroimaging and neuropsychology. *Brain*, *127*(4), 851-859.

Shohamy, D., Myers, C., Hopkins, R., Sage, J., & Gluck, M. (2009). Distinct hippocampal and basal ganglia contributions to probabilistic learning and reversal. *Journal of Cognitive Neuroscience*, *21*(9), 1820-1832.

Shohamy, D., Myers, C., Onlaor, S., & Gluck, M. (2004). Role of the Basal Ganglia in Category Learning: How Do Patients With Parkinson's Disease Learn? *Behavioral Neuroscience*, *118*(4), 676–686.

Shohamy, D., & Wagner, A. (2008). Integrating Memories in the Human Brain: Hippocampal-Midbrain Encoding of Overlapping Events. *Neuron*, *60*(2), 378–389.

Smith, L. B. (2000). Learning how to learn words: An associative crane. In R. M. Golinkoff et al. (Eds.), *Becoming a word learner: A debate on lexical*

acquisition. Oxford: Oxford University Press.

Smith, L. B., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568.

Smith, L. B., Yu, C., & Pereira, A. (2011). Not your mother's view: The dynamics of toddler visual experience. *Developmental Science*, *14*(1), 9-17.

Sobel, D., Tenenbaum, J., & Gopnik, A. (2004). Children's causal inferences from indirect evidence: Backwards blocking and bayesian reasoning in preschoolers. *Cognitive Science*, *28*(3), 303-333.

Song, S., Miller, K., & Abbott, L. (2000). Competitive hebbian learning through spike-timing-dependent synaptic plasticity. *Nature Reviews Neuroscience*, *3*(9), 919-926.

Spence, C., & Driver, J. (1997). Audiovisual links in exogenous covert spatial orienting. *Attention, Perception, & Psychophysics*, *59*(1), 1-22.

Spivey, M. J. (2007). *The continuity of mind*. Oxford: Oxford University Press.

Spivey, M. J., & Knowlton, S. (2008). Anti-heroism in the continuum of good and evil. In R. S. Rosenberg (Ed.), *The psychology of superheros: An unauthorized exploration* (p. 51-63). USA: Benbella Books.

Squire, L., Stark, C., & Clark, R. (2004). The Medial Temporal Lobe. *Annual Review of Neuroscience*, *27*, 279–306.

Squire, L., & Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science*, *253*(5026), 1380–1386.

Stager, C., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, *388*(6640), 381-382.

Sugita, Y. (2009). Innate face processing. *Current Opinion in Neurobiology*, *19*(1), 39-44.

Summerfield, C., Trittschuh, E., Monti, J., Mesulam, M., & Egner, T. (2008). Neu-

ral repetition suppression reflects fulfilled perceptual expectations. *Nature Neuroscience*, *11*(9), 1004-1006.

Suzuki, W. A. (2009). Perception and the medial temporal lobe: Evaluating the current evidence. *Neuron*, *61*(5), 657-666.

Swingley, D. (2008). The Roots of the Early Vocabulary in Infants' Learning From Speech. *Current Directions in Psychological Science*, *17*(5), 308–312.

Tanaka, J., & Curran, T. (2001). A neural basis for expert object recognition. *Psychological Science*, *12*(1), 43-47.

Teinonen, T., Fellman, V., Näätänen, R., Alku, P., & Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience*, *10*(1), 21-29.

Tenenbaum, J., Kemp, C., Griffiths, T., & Goodman, N. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*(6022), 1279-1285.

Thelen, E., & Bates, E. (2003). Connectionism and dynamic systems: are they really different? *Developmental Science*, *6*(4), 378–391.

Thiessen, E. (2007). The effect of distributional information on children's use of phonemic contrasts. *Journal of Memory and Language*, *56*(1), 16-34.

Thiessen, E., Hill, E., & Saffran, J. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, *7*(1), 53–71.

Toro, J., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition*, *97*(2), 25–34.

Tottenham, N., Tanaka, J., Leon, A., McCarry, T., Nurse, M., Hare, T., et al. (2009). The nimstim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, *168*(3), 242-249.

Tovee, M., Rolls, E., & Azzopardi, P. (1994). Translation invariance in the responses to faces of single neurons in the temporal visual cortical areas of

the alert macaque. *Journal of Neurophysiology*, *72*(3), 1049-1060.

Turk-Browne, N., Jungé, J., & Scholl, B. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology-General*, *134*(4), 552–563.

Turk-Browne, N., Scholl, B., Chun, M., & Johnson, M. K. (2009). Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience*, *21*(10), 1934-1945.

Turk-Browne, N., Scholl, B., Johnson, M., & Chun, M. (2010). Implicit perceptual anticipation triggered by statistical learning. *Journal of Neuroscience*, *30*(33), 11177-11187.

Turk-Browne, N., Yi, D., & Chun, M. (2006). Linking implicit and explicit memory: common encoding factors and shared representations. *Neuron*, *49*(6), 917–927.

Venables, W., & Ripley, B. (1999). *Modern applied statistics with s-plus, 2nd edition*. New York: Springer.

Wade, T., & Holt, L. L. (2005). Incidental categorization of spectrally complex non-invariant auditory stimuli in a computer game task. *Journal of Acoustical Society of America*, *118*, 2618-2633.

Wang, H., Hwang, A., & Pomplun, M. (2010). Object frequency and predictability effects on eye fixation durations in real-world scene viewing. *Journal of Eye Movement Research*, *3*(3), 1-10.

Waterfall, H. (2006). *A little change is a good thing: Feature theory, language acquisition and variation sets*. Unpublished doctoral dissertation, University of Chicago.

Werker, J., & Tees, R. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49-63.

Wertheimer, M. (1938). A source book of gestalt psychology. In W. Ellis (Ed.), (p. 71-88). New York, NY: Harcourt.

Wiesmann, M., & Ishai, A. (2010). Training facilitates object recognition in cubist paintings. *Frontiers in Human Neuroscience*, *4*, 1-7.

Winawer, J., Witthoft, N., Frank, M., Wu, L., Wade, A., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, *104*(19), 7780-7785.

Wu, R., Gopnik, A., Richardson, D. C., & Kirkham, N. (2011). Infants learn about objects from statistics and people. *Developmental Psychology*, *47*(5), 1220-1229.

Wu, R., & Kirkham, N. (2010). Visual statistical learning with and without an attention cue in infancy. *Journal of Vision*, *10*(7), 467-.

Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences*, *105*(13), 5012-5015.

Yoshida, K., Pons, F., Maye, J., & Werker, J. (2010). Distributional phonetic learning at 10 months of age. *Infancy*, *15*(4), 420-433.

Yu, C., & Smith, L. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, *18*(5), 414–420.

Yu, C., Smith, L. B., Christiansen, M., & Pereira, A. F. (2007). Two views of the world: Active vision in real-world interaction. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th annual confernce of the cognitive science society* (p. 731-736). Austin, TX: Cognitive Science Society.

Yuille, A., & Kersten, D. (2006). Vision as bayesian inference: analysis by synthesis? *Trends in Cognitive Sciences*, *10*(7), 301-308.